Chapter 3

Link Grammar and It’s Parsing with English

3.1 Introduction

Link Grammar is a formal grammatical system defined by Daniel D.K. Sleator and Davy Temperley in 1991 on the basis of natural language property which states that if arcs are drawn connecting each pair of words that relate to each other, then the arcs will not cross. The authors have written a link grammar of roughly seven hundred definitions that captures many phenomenon of English grammar and developed an algorithm for parsing sentences according to the given grammar. The system has no explicit notions of constituents or categories. Constituents are widely regarded as an essential component of any serious linguistic systems instead link grammar is word based- categories which play no role, yet which captures distributional regularities, handles recursion and is sufficiently constrained. In this formalism, a language is defined by a grammar that includes the words of the language and their linking requirements. A given sentence is accepted by the system if the linking requirements of all the words in the sentence are satisfied (connectivity), none of the links between the words cross each other (planarity) and there can exist at most one link between any pair of words (exclusion). A set of links between the words of a sentence that is accepted by the system is called a linkage. The grammar is defined in a dictionary file and each of the linking requirements of words is expressed in terms of connectors in the dictionary file. In the Section 3.2 basic definition and notions of Link Grammar is explained followed by detail explanation on how different phenomenon of English Link Grammar functions in Section 3.3. The algorithm developed especially for Link Grammar is discussed in Section 3.4 along with the special features used to fasten the parsing process followed by concluding remarks.
3.2 Basic Definition and Notions

A link grammar consists of a set of words like the terminal symbols of the grammar. A sequence of words is a sentence of the language defined by the grammar if there exists a way to draw arcs among the words. Each word has a linking requirement. The linking requirements of each word are contained in a dictionary. Following Figure 3.1 illustrates linking requirement of words* with a simple dictionary for the words ‘a’, ‘the’, ‘cat’, ‘snake’, ‘Mary’, ‘ran’ and ‘chased’.

Think each word with some connectors coming out. The connectors are either pointing left or right direction. A link between two words will be satisfied by connecting right pointing connector with left pointing connector of the same link type. Plugging a pair of connectors together corresponds to drawing a link between the pair of words. The Figure 3.2 shows how the linking requirements are satisfied in the sentence ‘The Cat Chased a Snake’.

The unused connectors stay as it is. Few more sentences are acceptable by the above simple dictionary like ‘A snake chased Mary’, ‘Mary ran’ etc. Thus the sequence of words: ‘the Mary chased cat’ is not acceptable by above dictionary shown in Figure 3.3 as it violates any one of the three rules defined.
Similarly ‘ran Mary’, ‘cat ran’ will also be not accepted. The three rules or conditions which have to be satisfied at the time of linking words is explained as follows,

- **Planarity** - When the links drawn above the words they should not cross each other.

- **Connectivity** - The links should suffice to connect all the words of the word sequence together.

- **Satisfaction** - The links should satisfy the linking requirements of each word in the word sequence.

A set of links that prove that a sequence of words is in the language of a link grammar is called a linkage. The link grammar dictionary consists of a collection of entries, each of which defines the linking requirements of one or more words. In a dictionary, the linking requirements are specified by means of a formula of connectors combined by the binary associative operators ‘&’ and ‘or’. Precedence is specified by means of parentheses. The connector is simply a character string ending in ‘+’ or ‘−’. Following Table 3.1 describes the dictionary entries

<table>
<thead>
<tr>
<th>Words</th>
<th>Formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td>a the</td>
<td>D+</td>
</tr>
<tr>
<td>snake cat</td>
<td>D− &amp; (O− or S+)</td>
</tr>
<tr>
<td>mary</td>
<td>O− or S+</td>
</tr>
<tr>
<td>ran</td>
<td>S−</td>
</tr>
<tr>
<td>chased</td>
<td>S− &amp; O+</td>
</tr>
</tbody>
</table>

Table 3.1: Dictionary Entries

The ‘+’ or ‘−’ suffix on a connector name indicates the direction (relative to the word being defined) in which the matching connector must lie. The ‘&’ of two formulas is satisfied by satisfying both the formulas. The ‘or’ of two formulas requires that exactly one of its formulas be satisfied. The order of the argument of an ‘&’ operator is significant.

Along with the rules like Planarity, Connectivity, Satisfaction and other rules like
Ordering and Exclusion are also necessary to get satisfied by each words formula which states

- **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 must satisfy a connector appearing to the left of that of the other word. Similarly a link to the right must satisfy a connector to the left of a longer link to the right.

- **Exclusion** - No two links may connect the same pair of words.

A different way of expressing a link grammar was introduced which is called Disjunctive Form. Basically it was introduced for resolving representational problem as it was cumbersome for mathematical analysis of link grammars, and in describing algorithms for parsing link grammars. 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 and the right list contains connectors that connect to the right of the current word. A disjunct will be denoted:

\[ ((L_1, L_2, \ldots, L_m), (R_n, R_{n-1}, \ldots, R_n)) \]

Where \( L_1, L_2, \ldots, L_m \) are the connectors that must connect to the left, and \( R_n, R_{n-1}, \ldots, R_n \) are connectors that must connect to the right. The number of connectors in either list may be zero. The trailing ‘+’ or ‘-’ may be omitted from the connector names when using disjunctive form, since the direction is implicit in the form of the disjunct. To satisfy the linking requirements of a word, one of its disjuncts must be satisfied (and no links may attach to any other disjunct). To satisfy a disjunct all of its connectors must be satisfied by appropriate links. The words to which \( L_1, L_2, \ldots, L_m \) linked, are to the left of the current word, and increasing monotonically in distance from the current word. The words to which \( R_n, R_{n-1}, \ldots, R_n \) are linked to the right of the current word, and are monotonically increasing in distance from the current word.

It is also easy to translate a formula into a set of disjuncts. This is done enumerating all ways that the formula can be satisfied. Following formula will be enumerated as shown in the Table 3.2 below,

\[ (A- \text{ or ( )}) \& \text{ D-} \& (B+ \text{ or ( )}) \& \text{ O- or S+}) \]

In a dictionary, connector name begins with one or more upper case letters followed by a sequence of lower case letters or *s. Each lower case letter is a subscript. Each
Table 3.2: Disjuncts for Given Formula

subscript provide various information (e.g. gender (m/f), number (s/p)) describing some properties of the underlying word. To determine if two connectors match, delete the trailing ‘+’ or ‘-’, and append an infinite sequence of *s to both connectors. The connector match if and only if these two strings match under the provision that * matches a lower case letter (or *). This dictionary allows certain connectors to be able to connect to one or more links. This makes it easy, for example, to allow any number of adjectives to attach to a noun, by putting ‘@’ before connector name is possible and called as multi-connector. The dictionary consists of a sequence of entries, each of which is a list of words separated by spaces, followed by a colon, followed by the formula defining the words, followed by a semicolon. If a word has more than one distinct meaning, it is useful to be able to give it two definitions. This is accomplished by defining several versions of the word with differing suffixes. The suffix always begins with a '.', which is followed by one or more characters. The convention indicated in this dictionary is; ‘.v’ indicates verb, ‘.n’ indicates noun and ‘.a’ indicates adjective etc.

3.3 The Working of the English Link Grammar

English is primarily a fixed word order language, i.e. the positions of various chunks in a sentence are usually fixed. English basically follows SVO i.e. Subject, Verb and Object word order. For example “The cat chased the mouse”, “Jon ate apple” etc. where subject comes first followed by verb and then object. But there are few cases when the word order in English sentence may deviate from their usual sequence. Some cases are given below,

- Object requirement of a verb is usually satisfied by some word to its right. However in some sentence types some word on the left of verb may act as its object. For example “The man the dog chased was old”. In this sentence the word order is OSV.
- Usually, adjective precede the noun they modify, but in some cases the adjectives
may follow the noun. For example, consider the sentences “This is the only available room” and “This is the only room available” in these sentences ‘available’ modifies the noun ‘room’.

- Adverbs may come before or after the verb they modify. For example, consider the position of the adverb ‘also’ in the sentences “I may also leave” and “I may leave also”.

- Adverb clause may precede or follow the sentence it modifies. Consider, “After he comes, I will go” and “I will go after he comes”. In both the sentences the adverb clause is ‘after he comes’, and it precedes the main clause in the first sentence and follows it in the second sentence.

By definition English link grammar fixes the functional role of the words to the left and to the right of a particular link. To achieve this links of different words are associated with ‘+’ or ‘-’ sign specifying whether that word makes the link with some word to its right or left, respectively. Since the cases when the word order is not fixed are very few, separate links are defined to be used in those cases.

### 3.3.1 Noun Phrase Links

English noun phrase consists of a noun/pronoun, called the head or antecedent of the noun phrase, and any of the following optional constituents: determiners, pre-nominal modifiers and post-nominal modifiers. Various components and the corresponding links used in English noun phrases are discussed below. It may be noted that English modifiers do not inflect.

#### 3.3.1.1 Determiners

Determiners are used very prominently in English to specify the definite or indefinite aspect of the antecedent. English link grammar defines various links for determiners, for example nouns have a D- connector and determiners have D+ connector. On singular nouns, the D- is mandatory; on plural nouns, it is optional. Some nouns, like ‘war’, can be either singular or mass; here the D- is optional. Some determiners can be used with any noun, like ‘the’ and ‘his’. Others can be only used with singular nouns, like ‘a’. Others can only be used with mass or plural nouns, like ‘all’ or ‘most’. Others like ‘many’ can only be used with plural nouns. Subscripts are used to enforce these rules. DD connector is used to connect the definite determiners with the number expressions and adjectives acting as nouns.

DG link connects the word ‘the’ with proper nouns (e.g. ‘The Napoleon’). DP link connects possessive determiners to gerunds in case where gerund is taking its
normal complement. DT link connects determiner with nouns in certain idiomatic time expression ‘last week’ or ‘next Monday’ etc.

3.3.1.2 Pre-nominal Modifier

Word(s) which modifies the noun and which appears before the noun are the Pre-nominal modifiers. Usually they come after the determiners. Pre-nominal modifiers can be of three types: Adjective, e.g. big building, participle (present or past), e.g. ‘singing girl’, ‘deserted village’; and noun, e.g. ‘lady doctor’. To connect such type of pre-nominal modifiers with head noun various links are used like ‘A’ link is used to connect attributive adjectives with noun and participle modifiers; so adjective have an A+ and nouns have an @A-. The sign @ before any link type allows the connector to link to one or more other connectors for e.g. “the big black ugly dog chased John” here “the big black ugly dog” is a noun phrase and adjectives big black and ugly are connected with noun dog with the link ‘A’.

Link ‘AN’ is used to connect the noun modifier with the noun, for e.g. “The tax proposal was rejected”, in this sentence tax is the noun which is modifying the head noun of a sentence and link AN will connect them.

3.3.1.3 Post Nominal Modifier

In English sentences post nominal modifiers are used in many forms, such as: Adjective, e.g. ‘the President designate’; Prepositional phrase, e.g. ‘the girl in the room”; gerund e.g. ‘the girl singing in the rain”; past participle, eg. ‘the train bound for delhi’. In such cases ‘M’ link, with various suffixes viz. ‘a’ for adjectival modifier, ‘p’ for prepositional phrases modifying nouns, ‘g’ and ‘v’ for participle modifiers, ‘r’ for possessive relatives etc. is used. In some cases noun phrase is surrounded by commas, consisting of a phrase like ‘many [/some/five/several] of them’ followed by a noun phrase. Here ‘of them’ is treated as idiomatic expression which serves as the head of the phrase and ‘many [/some/five/several]’ gets connected with link MF. Similarly MG link allows certain prepositions to modify proper nouns. MX link connects nouns to post nominal noun modifiers surrounded by commas.

3.3.2 Noun to Verb Links

As we know in the vast majority of cases, the central clause elements Subjects (S), predicative verbs (V), objects (O) and predicative (P) occur in a fixed order. The main rule which is followed is the subject precedes the predicate verb. S is the link which connects subject/noun to finite verb. All nouns carry S+ connector. Many other words like nominative pronouns, determiners like ‘many’, ‘all’, ‘this’, Miscellaneous words
like ‘someone’, ‘whatever’, ‘everything’; numbers; possessives like ‘hers’ and ‘mine’; gerunds; and adjectives carry S+ link as well. The basic subject-verb agreement rule in English is very simple. It states that a singular subject takes a singular verb, while a plural subject takes a plural. However this basic rule is not always true with various types of verbs and hence links which joins this agreement are neatly crafted. Link $S_s$ connects singular nouns words to singular verb forms; $S_p$ connects plural nouns to plural verb forms. Proper nouns carry $S_s$ link and thus are assumed to be singular. Simple past verb forms do not distinguish between singular and plural thus an unsubscripted $S$ link is used. Some special S links are used like $S_{sgt}$ link is used for a few nouns that can take ‘be+that’ as predicates for example in sentence ‘the idea was that we would go to London’. Link $S_{sbt}$ is used for a few subjects ‘it’, ‘this’, ‘that’ that can take the predicate ‘because’: ‘this is because he is stupid’. $S_{sqt}$ is used for subjects like ‘questions’ that can take predicate ‘indirect question’ : ‘The question is why he did it’. $S_{sgt}$ is used for gerunds: ‘-ing’ forms of verbs that act like nouns or head of noun phrases. Use of gerunds is extremely problematic and hence by observing rules of governing in actual construction and context links are designed and used appropriately. SF is a special subject link type used for certain filler subject like ‘it’ and ‘there’. This interacts heavily with post-processing. SFI connects filler subjects ‘it’ and ‘there’ to invertible verbs in questions with S-V inversion for example ‘Is it like that Jane will go’. Link SI is used in Subject-Verb inversion eg ‘Is Joe coming’. All words carrying S+ nouns, nominative pronouns and many other words carry SI as well. SI is used for singular noun and verb forms. $SI_p$ is used for plural forms. SX is used to connect the first person pronoun I to the verbs ‘was’ and ‘am’ and SXI used in case of S-V inversion.

### 3.3.3 Adjective Links

Adjectives usually appear in noun phrase preceding to head noun hence Link A connects pronoun adjective to noun. There are few variations in adjective in their appearance like in the cases where the adjective is fronted such as question and indirect questions and then Link AF connects adjective to verbs for example ‘How big is it’ or ‘I wonder how big he wants it to be’. LE is a special connector used in comparative construction to connect an adjective to the second half of the comparative expression beyond a complement phrase for example in sentence ‘It is more likely that Joe will go than that Fred will go’ here adjective likely gets connected to word ‘than’ with LE link. TA link is used to connect adjective like ‘late’ or ‘early’ to month names for example ‘We did it in early December’. TD link connects day of the week words forward to words like ‘morning’, ‘afternoon’ and ‘evening’.

28
3.3.4 Preposition Links

Several prepositions exists in English grammar like in, on, into, onto, under, above, at etc. generally a prepositions come before a noun to show the noun’s relationship to another word in the sentence. Connector J is used to connect prepositions which have J+ link to their objects, proper and common nouns, accusative pronouns and other words that can act as noun phrase have J- connector.

JG is the connector which connects certain prepositions (of and for) to proper noun objects for example “The National Academy of Science”. JQ link is used to connect prepositions to question words like “In which room were you working”. The preposition ‘from’ to various other prepositions is connected through link FM.

3.3.5 Relative Clause

Relative clause is used to join two English sentences or to give more information about something. There are actually two types of relative clause first is defining and the other is non-defining relative clause. There are various relative pronouns like ‘who’, ‘which’, ‘whose’, ‘whom’, ‘that’ used as subject or object pronoun. Subject or Object pronouns cannot be distinguished by their forms, however they can be distinguished by a simple rule “If the relative pronoun is followed by a verb, the relative pronoun is subject pronoun” and “if the relative pronoun is not followed by a verb but by a noun or pronoun, the relative pronoun is an object pronoun. In link grammar formalism R link is used to connect nouns to relative clauses. In subject-type relatives, it connects to the relative pronoun. In object-type relatives, it connects either to relative pronoun or the subject of the relative clause if the relative pronoun is omitted. Link RS is also used in support with R link which is used in subject-type relative clauses. Eg “The dog who chased me was black”.

3.3.6 Subordinate Clause Links

Subordinate clause is a group of words that has both a subject and a verb but cannot stand alone as a sentence. Finite clauses are introduced by a subordinating conjunction which introduces the subordinate clause. There are various groups of conjunctions like simple conjunctions for eg ‘when’, ‘whenever’, ‘where’, ‘whenever’, ‘because’, ‘if’ etc, conjunctive groups for eg. ‘As if’, ‘as though’, ‘even if’, ‘soon after’ etc. and there are some complex conjunctions also exists for eg provided, considering, suppose, in case, to the extent that, as soon as etc. In link grammar formalism link C connects subordinating conjunctions and certain verbs and adjectives with subjects of clauses. It is therefore used only in embedded and subordinate clauses, not main clauses. Eg “Call me when you are ready” CO is one of link which is used to connect “openers”
to subjects of clauses. Eg “Apparently they went to movie, On Tuesday they went to movie”. CC is a link used mainly to connect clauses to coordinating conjunctions. It is used with coordinating conjunctions only and it links to the subjects of the previous main clause.

### 3.3.7 Object Related Links

In English grammar an object in a sentence is the entity that acts upon by the subject. Object is taken as part of the predicate. The main verb in a clause determines if and what objects are present. Transitive verbs require the presence of an object whereas intransitive verbs block the appearance of an object. Various object types are commonly acknowledged: direct, indirect and prepositional. Direct objects is the entity acted upon for eg “The girl ate fruit” here fruit is direct object. The indirect object is entity indirectly affected by the action for eg “She sent him a present” here him is the indirect object where as prepositional object is an object introduced by a preposition. To connect object either direct or indirect with transitive verbs link O is used. Words carrying O- includes nouns, pronouns and many other kinds of words that can act as noun phrases or heads of noun phrases. Transitive verbs carry O+. This is conjoined with other left pointing connectors which link to the subject such as S-, PP-, I-, Pg-. Os- and Op- connectors mark nouns as being singular or plural.

### 3.3.8 Verb Links

Generally the word order of English grammar is S-V-O i.e Subject-Verb-Object. As per link grammar formalism we have already discussed that, this word order be preserved by its rules i.e. Planarity, Connectivity and Satisfaction. When verb is connected with subject or object it has S or O link respectively. Simple past forms of verbs simply have an unsubscripted S- and transitive verbs have an O+ connector. Hence verbs do not possess separate links. Some variations in the verb are discussed below. Infinitive forms of verbs have an I- link. Most infinitive verb forms are the same as the simple plural form, so a single dictionary entry serves both of these functions. Verbs which take infinitives eg do, may, can etc have an I+ connector. They also have an S- connector. Past participle forms of verbs have a T- connector.

### 3.3.9 Question Links

Link Grammar handles many different kind of questions which involves ‘Wh’ questions like Who, Where, What etc. Link Q is used in questions, in several different ways. It is used to connect to auxiliaries in simple S-V inversion questions, when there is no preceding object. In some cases the auxiliary connects to the wall. In other cases it
connects to a question word like ‘when’, ‘where’ or ‘why’. Who question is dealt in two ways one is Who Subject type questions and the other is Who Object type question. In Who- Subject question, who simply acts as subject of the sentence. Simply S+ connector is given to Who ored with relative clause connectors. In Who object type question, subject and verb are inverted. To accomplish this SI connector is used in Conjunction with Q Connector. Post processing ensures that when a question word is used in Object type or when-where-why questions, the outer group of the sentence contains some kind of SI; this is because in all such cases a Wq is used to connect to the Wall and this requires on SI in the same group.

### 3.3.10 Adverb Links

The simple definition of Adverb is the word which modifies verb. Link E is used for verb-modifying adverbs which precedes the verb. There are different types of adverbs, some kind of adverb are quite specific in the way they may be used. They are highly constrained like adjective modifiers for this Link EA is used which connects adverbs to adjectives, adverb modifiers which connects adverbs to other adverbs Link EE is used for this modifiers, comparative modifiers: link EC which connects adverb and comparative adjectives, number-modifiers: link EN is used to connect certain adverbs to expression of quantity, modifiers of ‘as’ for this link EZ is used to connect certain adverbs to the word ‘as’, like ‘almost’, ‘just’ and ‘about’; and modifiers of prepositional phrases for which link MV1 is used.

Other adverbs are less constrained in the way they can be used: adverbs of manner, clausal, and time adverbs. Adverbs of manner refer specifically to the manner in which an action is done: ‘She speaks sweetly’. Clausal adverbs refer in some way to the clause as a whole ‘Ram is coming’. Time adverbs give information about the time of the action: ‘Sita is soon coming’. Each of these types of adverbs has a variety of syntactic usages; they also share some usages, and therefore have some connects in common. However the usage of each category is slightly different; there is also some variance within the categories.

### 3.3.11 Conjunction Links

Complete clauses have CL- connector, which can be used to connect verbs that take clauses. Nouns and pronouns have CL- connector & ed with the subject connectors. A conjunction such as because has an EV- and a CL+. That is, it links to the verb of the clause before it and it links to the subject of the clause after it. A conjunction can also begin the sentence, followed by a subordinate clause, followed by the main clause. As before, the conjunction links to the CL- connector on the subject of the subordinate clause. It also links to a CO- connector on the subject of the main clause.
The conjunction also has XC+ connector; commas have an XC- connector. ‘John says that after the party, Joe was angry’

3.3.12 Number Expression Links

In link grammar number expressions are handled separately. The Link names start with letter ‘N’ to indicate it is number expression followed by a letter which provides some information about the link type. Link ND connects number with certain expressions which require numerical determiners for eg. ‘I saw him THREE WEEKS ago’, where as Link NS connects singular number s like one, 1 to idiomatic expression. Link NW is used in case of idiomatic fraction expression. Link NF is used in idiomatic number expressions involving ‘of’. This NF link is used with link NJ for eg. ‘He lives two Thirds of a mile from here’ Link NN connects number words together in series for eg ‘Four Hundred Thousand Million People’. Link NR connects fraction words with superlatives for eg. ‘Third biggest city’.

3.3.13 Time Expression Links

For Time expressions separate links are proposed similar to number expression. Link TM is used to connect month names to day numbers. Link TY is used for certain idiomatic usage of year numbers. Link TW connects day of the week to month name I sentence like ‘I saw her on Wednesday, May 27, 1980’ Link ON is used to connect preposition ‘on’ to such time expressions. Link Y is also used in certain idiomatic time and place expression. It connects quantity expression to the head word of the expression for eg ‘We swam three miles away’.

3.3.14 Comparative Links

As per usage of comparative word few links are in use. Link CX is used in comparative construction as shown below ‘he has more money than I do’. Link AM is used in the comparative construction ‘as much’ and ‘as many’. Link ER is used in the construction ‘the X-er … the Y-er’ i.e. ‘The better it is. The more people will use it’. Link CQ connects to auxiliaries in comparative construction involving s-v inversion for eg ‘She has more money THAN DOES Joe’.

3.3.15 Miscellaneous Links

We found few more links which do not falls in any categories discussed above. Few of such links are discussed below, Link EL connects certain words like someone/everyone/no one/ anyone / somebody / somewhere / who / what/ where / whatever / whoever to
the word ‘else’ Link FL connects word ‘for’ to ‘long’ for eg I didn’t wait FOR LONG / FOR VERY LONG” Link NO given to words such as ‘um’ and ‘oh’ which have no grammatical meaning. Link NT connects word ‘not’ with ‘to’ for eg ‘I told him NOT TO go’ Link OF connects certain verbs and adjectives to the word ‘of’ for eg ‘I thought of something’ Link TH connects words that take ‘that’ complements with the word ‘that’. These include verbs, nouns and adjectives for eg ‘I assured her that’, ‘the idea that we’, ‘I’m certain that she’ Link TI is used for titles like ‘President’ and ‘Chairman’ which can be used in certain circumstances without a determiner; after the preposition ‘as’ or after verbs like name or elect for eg ‘As President of the company’, ‘She was named President of the company’. Link X is used to connect punctuation symbol to words. Link Xc connects a word to a comma to the right; Xd connects a word to a comma to a left. Xp and Xi are used for periods; Xx is used with colons and semi-colons. Link Yp is used in possessive constructions to connect plural noun forms ending in ‘s’ to apostopy “’”, for eg ‘the student’s wishes have been neglected’. Link Ys connects nouns to the possessive suffix ‘s’. ‘s’ then act as a determiner, making a D connection to a noun, for eg ‘Jane’s dog is black’

3.3.16 The Wall

The concept wall is introduced to solve the problem of question inversion which take place in relative clause. The wall can best be thought of as an invisible word which the program inserts at the beginning of every sentence and which has a connector expression like any other word. The wall has only optional connector; it does not have to connect to anything in the sentence. Link W is used to attach main clauses to the wall (left wall). Almost all kinds of main clauses, declaratives, most questions (except yes-no questions) and imperatives. The link RW connects the right hand wall in cases where the right hand wall is not needed for punctuation purposes.

3.4 The Parsing Algorithm

Researchers Daniel D. K. Sleator and Davy Temperley introduced basic parsing algorithm through the technical report [Sleator and Temperley, 1991] in October 1991 then in August, 1995 Dennis Grinberg, John Lafferty and Daniel Sleator came with robust parsing algorithm for link grammar [Grinberg et al., 1995] and they tested this ideas by parsing Switchboard corpus of conversational English. We will be discussing basic parsing algorithms in the following section.

The Algorithm Let us call it as the basic algorithm as it was introduced with the basic English grammar formalism. This algorithm for parsing link grammar is based on dynamic programming. It tries to build up a linkage in a top down
fashion. We have already discussed about a disjunct in Section 3.2. A disjunct d has pointers to two linked lists of connectors. These pointers are denoted left[d] and right[d]. If c is a connector then next[c] will denote the next connector after c in its list. The next field of the last pointer of a list has the value NIL. For example, suppose the disjunct d = ( (D,O) ( ) ). Then left[d] would point to connector O. next[left[d]] would point to connector D. next[next[left[d]]] would be NIL. Similarly right[d] = NIL. Suppose a sentence is given as input to algorithm. The words of the sequence to be parsed are numbered from 0 to (N-1). Let leftmost word be called L and rightmost word be called R. Consider the situation after a link has been proposed between a connector l’ on word L and a connector r’ on word R. For convenience let l be the next[l’] and r be the next[r’],

Figure 3.4: An Attempt of Forming a Link between Word L and R

The region will be denoted (L, · · · , R). Different cases handled are as follows, Case 1) If there are no words in this region. i. e. L = R+1 solution built is certainly invalid. If either l not equal to NIL or r not equal to NIL If l = r = NIL then region is OK.

Case 2) The region between L and R contains at least one word. To follow the rules of link grammar formalism the words of this region to the rest of the sentence there must be at least one link either from L to some word in this region, or from R to some word in this region.

Since the connector l’ has already been used in the solution being constructed, this solution must use the rest of the connectors of the disjunct in which l’ resides. The same holds for r’. The only connector l’ of these disjuncts that can be involved in the (L, · · · , R) region are those in the lists beginning with l and r. In fact, all of the connectors of these lists must be used in this region in order to have a satisfactory solution. Suppose for the moment that l is not NIL. This connector must link to some disjunct on some word in the region (L, · · · , R). The algorithm tries all possible such words and disjuncts. Suppose it finds a word w and a disjunct d on w such that the connection l matches left[d]. This link can be added into partial solution.
To extend this partial solution to a full solution two regions needs to examined region \((L, \cdots, W)\) and region \((W, \cdots, R)\). This is the basic functioning of an algorithm where links are searched in a region. The formal description of the algorithm is as given below,

1. **Function PARSE**
   
   (a) \(t \leftarrow 0\)
   
   (b) **for** each disjunct \(d\) of word \(0\)
   
   (c) \(\text{do if } \text{left}[d] = \text{NIL}\)
   
   (d) \(\text{then } t \leftarrow t + \text{Count}(0, N, \text{right}[d], \text{NIL})\)
   
   (e) \(t \leftarrow t + \text{Count}(0, N, \text{NIL}, \text{NIL})\)
   
   (f) **return** \(t\)

2. **Function COUNT** \((L, R, L', R')\)

   (a) **if** \(R = L+1\)
   
   (b) \(\text{then if } l = \text{NIL} \text{ and } r = \text{NIL}\)
   
   (c) \(\text{then return } 1\)
   
   (d) \(\text{else return } 0\)
   
   (e) **elseif** \(l = \text{NIL} \text{ and } r = \text{NIL}\)
   
   (f) \(\text{then total} \leftarrow 0\)
   
   (g) **for** each disjunct \(d\) of word \(L+1\)
   
   (h) \(\text{do if } \text{left}[d] = \text{NIL}\)
   
   (i) \(\text{then total} \leftarrow \text{total} + \text{Count}(L+1, R, \text{right}[d], \text{NIL})\)
   
   (j) \(\text{total} \leftarrow \text{total} + \text{Count}(L+1, R, \text{NIL}, \text{NIL})\)
   
   (k) **return** total
   
   (l) **else**
The function COUNT takes as input indices of two words, and a pair of pointers to lists of connectors. COUNT return a number. This is the number of different ways to draw links among the connectors in the lists pointed by l and r such that these links satisfy the following conditions:

1. They are planar (i.e. they do not cross). And they also satisfy the ordering and exclusion meta rules.
2. There is no link from a connector of the l list to one of the r list.
3. The requirement of each word strictly between L and R are satisfied by the chosen links.
4. All the connectors in the lists pointed to by l and r are satisfied.
5. The links either connect all the words \([L, \ldots, R]\) together, or they form two connected components: \([L, \cdot, M]\) and \([M + 1, \cdot, R]\) (for some M satisfying \(L \leq M \geq R\))

### 3.5 Concluding Remarks

The study of Link Grammar Formalism and its parsing with English includes introduction to the Link Grammar formalism which we have chosen as a framework for Marathi language. The chapter gives whole framework of parsing English language. The detail study of formulation of English language in the Link Grammar formalism gives the insight about how different linguistic phenomenon has adopted in this framework. We studied each link type and categorized them into various broad categories like noun phrase links, noun to verb links, adjective links and so on. This categorization will help new comers in Link Grammar formalism to understand it in minimum time.

Study of parsing algorithm is used in framing strategies in Marathi Link Grammar parsing. We shall be utilizing whole study included in this chapter to design and develop Marathi link types and algorithm suitable to parse Marathi Link Grammar to be formulated through it.