Formulation and Computer Program for Arrangements of Verbs with Subjects

H.S.Dhami
Director ICT,
Kumaun University
Nainital

Shalini Sharma
Project Fellow
Dept. of Mathematics
SSJ Campus Almora

Amit Kumar
Assistant Professor
Dept. of Mathematics
M.B.P.G.College Haldwani

ABSTRACT
The basic parts of a sentence are the subjects and verbs. Although many other structures can be added to this basic unit, the pattern of subjects plus verbs can be found in even the longest and most complicated structures. In this paper, we have made an attempt to address this problem by formulating mathematically the English Grammar rules for association of verbs with subjects so that a computer program could be generated.

KEYWORDS
Membership function, fuzzification of the subject, connectives, influence of common phrases on number of verb, collective noun, proper noun.

INTRODUCTION

Language has fascinated people for thousands of years and linguists have studied every detail, from the number of languages spoken in New Guinea to what we say razzle-dazzle instead of dazzle-razzle. The ingredients of language are words and rules. Words in the sense of memorized links between sound and meaning; rules in the sense of operations that assemble the words into combinations whose meaning can be computed from the meaning of the words and the way they are arranged.

The word "grammar" has many meanings. For some people, grammar specifies the "correct" way to speak or write. For others, the word refers to the inflections (the word endings) common in many languages. For still others, the grammar is about how humans organize ideas into words. The word "grammar" means all of those things, but for us, the word means something quite specific: how we choose and arrange our words in a sentence. We can consider it as an example of a combinatorial system, in which a small inventory of elements can be assembled by rules into an immense set of distinct objects.

Chomsky in his pioneering work not only laid the foundation of transformational grammar but also offered the promising computational approach to grammar by publishing a booklet "Syntactic structures" long back in 1957. The teacher of Chomsky, Zellig Sabbettai Harris (1982) has served the cause of scientific study of language by doing comprehensive study of English Grammar on mathematical principles incorporating semantic, syntactic and morphological details of natural language, as exemplified by English. Lambek (2004) has applied computational algebraic approach for framing constraints and transformations of English Grammar in the algebraic context. Akshat Kumar & Shivshankar B.Nair (2007) have provided a novel corpus based approach for grammar checking that uses the principles of an artificial immune system. Comparatively newer approach to grammar is mathematical, which belongs to the tradition of categorical grammar. In this approach, initiated by J.Lambek (2008), the principal idea is to assign types (elements of an algebraic system or terms of a logical system) to each word. The system of types has been called the free pre-group generated by the partially ordered sets of the basic types. Claudia Casadio and Joachim Lambek (eds.)(2008) have compiled the works of different researchers dealing with pre group grammars and Tupled pre group grammas related with different languages. Formal syntax framework for natural languages, particularly English, has been presented by Yingxu Wang (2009). In another work, Yingxu Wang (2010) has used denotational mathematics in brain Informatics. Pawet Scheffler and Marcin Cinciate (2010) in an article have reported an empirical study that examines to what extent learners can identify and understand the grammatical structure they produce when they speak spontaneously.

We are embarking upon explaining mathematically the aspect of association of verbs with subjects in English language using the concept of fuzzy logic. The computer program of the process in visual basic as front end and C++ as back end is also being given.

PRELIMINARIES

Usage of verb in sentences depends on the subject .The factors contributing to the agreement of verb with the subject are the **person** of the subject , its **number** , its usage as a **collective noun or a proper noun** , use of various **connectives** like **'and', 'or'** etc.

The basic structure of the sentence can be formulated as

<Article/ Noun/Pronoun> <Subject(S_1)> <Connective> <Subject(S_2)>

<Verb+remaining object>

There can be one subject or more than one subjects joined by some connective, which can be categorized into three classes, as under:

Class I : and

Class II : "or" , "nor", "either … or", "neither …nor.

Class III : "with", "together with ", "in addition to", "as well as"

Thus the part of the sentence under consideration for studying agreement of the verb with

the subject is <S₁> <CONNECTIVE> < S₂> <VERB>

As the usage of verb depends on the number and person, which further depends on such factors that show gradual variations and cannot be expressed using crisp sets falling under the purview of abstract algebra, so the use of Fuzzy sets has been made to explain the rules

MODUS OPERANDI

The association of verb with the subject has been discussed in two stages:

(I) Analysis and fuzzification of subject on the basis of person and number.

(II) Agreement of the verb for the three defined classes on the basis of above analysis.

Step I: Analysis and fuzzification of subject on the basis of person and number.

Agreement of the verb with the subject (S) on the basis of **person** can be mathematically formulated as follows:

I   A membership function φ is defined to assign a value to the **person** of the subject as under:

$$\varphi = \left\{\frac{n}{3} : n = 0,1,2\right\} \ldots\ldots\ldots.(1.1)$$

embodying the conditions-

(i)   If the subject is in the third person, that is if n=0, $\varphi(S) = 0$

(ii)   If the subject is in second person, that is, if n=1,

$$\varphi(S) = \frac{1}{3}$$

(iii)   If the subject is in first person, that is if n=2 ; $\varphi(S) = \frac{2}{3}$

Association of the verb with the subject in **number,** depending on the person of the subject(s) , usage of subject(s) in a phrase,

influence of articles or pronouns before each subject , influence of words such as **each , every , many, a , one** before the subject can be mathematically formulated on the basis of fuzzy sets as follows:

Let ξ be a membership function defined as "**number of the subject**"

$$\xi = \left\{\frac{n}{4} : n = 0,1,2,3\right\} \ldots\ldots\ldots\ldots(1.2), then$$

(i)   If S_1, S_2 are singular subjects then $n = 0$, $\xi(S) = 0$

(ii)   If there is only one subject which is in plural form but singular in meaning

or S is a collective noun and when the collection is thought as a whole then $n=1$,

$$\xi(S) = \frac{1}{4}$$

(iii)   If there is only one subject which is singular in form but plural in meaning

or the subject is a collective noun such that, the individuals of which, it is   composed fall in

the category $n=2$ and so $\xi(S) = \frac{1}{2}$

(iv)   If any one of the subjects is plural, then n = 4 ,$\xi(S) = 1$

The membership function η for "**persons**" influencing the number of the verb can be defined as under:

$$\eta = \{1 - x : x = 0,1\} \ldots\ldots\ldots.(1.3)$$

with the conditions that

(i)   If both the subjects are of the same person or there is one subject only, then.

x=1, $\eta(S) = 0$

(ii)   If both the subjects are not of the same person, then

x=0 , $\eta(S) = 1$

The membership function U for the" **influence of common phrases on number of verb"** can be defined as**:**

$$U = \left\{\tfrac{n-1}{2} : n = 1,2\right\} \ldots\ldots(1.4)$$

with the restrictions:

(a) If the subjects along with the conjunction "**and**" form a common phrase then $n=1$, $U(S) = 0$

(b) If the subjects along with the conjunction "**and**" do not form a common phrase then $n = 2$, $U(S) = \dfrac{1}{2}$

The presence of noun or pronoun before both the subjects influences the number of the verb. We have defined the **influence of articles or pronouns before each subject on the number of the verb** in terms of membership function V as follows:

$$V = \left\{\tfrac{2y-1}{4} : y = 1,2\right\} \ldots\ldots(1.5),\text{ with the conditions that:}$$

(a) If an **article or pronoun** is before both the subjects then $y = 2$, $V(S) = \dfrac{3}{4}$

(b) If an **article or pronoun** is not before both the subjects or there is one subject only, then $y = 1$, $V(S) = \dfrac{1}{4}$

The membership function W designed for defining the influence of **each , every , many a , one, of** on the number of the verb can be represented as follows:

$$W = \left\{\tfrac{2n}{5} : n = 0,2\right\} \ldots\ldots(1.6),\text{ with the proviso that}$$

(a) If the subject(s) is(are) preceded by **each or every , many a , one of ,** then
$n = 0$, $W(S) = \dfrac{0}{5} = 0$

(b) If the subject(s) is(are) not preceded by **each or every, , many a , one of ,** or

there is one subject only, then $W(S) = \dfrac{4}{5}$

Step II: Agreement of the verb for the three defined classes on the basis of above analysis.

The degree of number of the verb may be defined in terms of a function X as:

$$X = \xi\,(S) + \eta\,(S) + U(S) + V(S) + W(S) \ldots\ldots(1.7)$$

Following two cases may arise:

(a) If $X > 2$, then plural form of verb shall be used.

(b) If $X \leq 2$, then singular form of verb shall be used.

Same set of conditions shall be applicable if only one subject is present.

Now let Y be a membership function defining **person** of verb with the subject and as such Y will be a function of S_1 and S_2, defined as:

$$Y = \varphi\,(S_1 \cup S_2) = \max\{\varphi\,(S_1) , \varphi\,(S_1)\} \ldots(1.8)$$

The different cases arising in this situation can be formulated mathematically as:

1. The verb shall be in agreement with the first person, if
$$Y = \dfrac{2}{3} \quad\ldots\ldots\ldots(1.9)$$

2. The Verb shall be in agreement with the second person, if
$$Y = \dfrac{1}{3} \quad\ldots\ldots\ldots(1.10)$$

3. The Verb shall be in agreement with the third person, if
$$Y = 0 \quad\ldots\ldots\ldots(1.11)$$

If the connectives are:

"**or**" , "**nor**" ,"**either…or**" ,"**neither…nor**" , then X shall be a function of

$\xi\,(S_1)$ **and** $\xi\,(S_2)$, defined as

$$X = \xi\,(S_1 \cup S_2) = \max\{\,\xi(S_1), \xi(S_2)\,\} \ldots\ldots(1.12)$$

The mathematical conversion of the arising possibilities, that is,

If singular form of verb is used, then

$$0 \leq X \leq \frac{1}{4} \quad \text{........(1.13)}$$

and if either plural form of verb is used or the subjects are placed in such a way that

$$\xi\,(S_2) \geq \xi\,(S_1), \text{ then}$$

$$\frac{1}{4} \leq X \leq 1 \quad \text{........(1.14).}$$

Hence Y may be expressed as a function of S_2 and shall be expressible as:

$$Y = \varphi\,(S_2) \quad \text{.........(1.15)}$$

The probable cases for different values of Y have been discussed by the equations (1.9), (1.10) and (1.11).

If connective is "**with**" , "**together with**", "**in addition to**", "**as well as**" , then X may be considered as a function of $\xi\,(S_1),$ defined as

$$X = \xi\,(S_1)\text{............(1.16)}$$

Here again the same type of cases shall arise as discussed above in (1.13) and (1.14).

Therefore Y shall be a function of ξ (S_1), defined as:

$$Y = \varphi\,(S_1)\text{.........(1.17)}$$

The probable cases for different values of Y shall be same as those discussed by equations (1.9), (1.10) and (1.11). Same situation shall arise if no connective is present.

From the above discussion it is clear that the number and person of the verb is governed by two parameters X and Y and the arrangements for nature of verb can be summarized as under:

If the connective is "**and", then** X can be obtained in 2 ways and Y can be obtained in 3 ways. Hence total number of ways in which verb can be obtained shall be $2 \times 3 = 6$. This fact can be understood with the help of following table:

	X≤ 2	X > 2
Y=$\frac{2}{3}$	Singular	Plural
	First person	First person
Y=$\frac{1}{3}$	Singular	Plural

	Second person	Second person
Y=0	Singular	Plural
	Third person	Third person

If the connective is any one of

"**or**" , "**nor**" ,"**either…or**" ,"**neither…nor**" ,"**with**" ,"**together with**", "**in addition to**", "**as well as**" , then

X can be obtained in 2 ways and Y in 3 ways. Hence total number of ways in which verb can be obtained shall be $2 \times 3 = 6$. The process can be explained as depicted in the table given as under:

	$0 \leq X \leq \frac{1}{4}$	$\frac{1}{4} \leq X \leq 1$
Y=$\frac{2}{3}$	Singular	Plural
	First person	First person
Y=$\frac{1}{3}$	Singular	Plural
	Second person	Second person
Y=0	Singular	Plural
	Third person	Third person

Following examples shall illustrate the designed rules.

1. The captain and the adjutant **are** present.

   Since the connective is **and** which belongs to Class I (as discussed by us).

   Also $\quad X = \xi\,(S) + \eta\,(S) + U(S) + V(S) + W(S)$
   ……from(1.7)

   $$Y = \varphi\,(S_1 \cup S_2) \;=\; \max\{\varphi\,(S_1)\,,\,\varphi\,(S_1)\}$$
   ……from(1.8)

   As both the subjects are singular, so $\xi\,(S) = 0$

   According to the condition (b) of equation (1.4)

   $$U(S) = \frac{1}{2}$$

   The value of V(S), obtained according to the proviso (b) of equation (3.6) shall be

   $$V(S) = \frac{3}{4}$$

and according to condition (b) of (1.6)

$$W(S) = \frac{4}{5} \text{ , so}$$

$$X = 0 + \frac{1}{2} + \frac{3}{4} + \frac{4}{5} + 0 = \frac{41}{20} > 2$$

Hence the plural form of verb will be used.

Since $\varphi(S_1), \varphi(S_2) = 0$

and

$Y = \varphi(S_1 \cup S_2) = max\{0,0\} = 0$, so $Y = 0$

Hence the verb is in agreement with the third person as discussed in expression (3.11), which accentuates that Y will remain unaltered.

Thus the plural form of verb in third person (are) has been used.

2. Neither he nor I was there.

Since the connective is **Neither ….. nor** which belongs to Class II (discussed earlier).

$$X = \xi (S_1 \cup S_2) = max\{ \xi(S_1), \xi(S_2) \}$$

……..from(1.12)

$Y = \phi (S_2)$ ………from (1.15)

As both the subjects are singular; $\xi (S) = 0$

$$X = \xi (S_1 \cup S_2) = max\{ \xi(S_1), \xi(S_2) \} = max\{ 0,0 \} = 0 < 2$$

Hence the singular form of verb will be used.

S_1 is third person so $\varphi(S_1) = 0$

S_2 is first person so $\varphi (S_2) = \frac{2}{3}$

$$Y = \varphi (S_2) = \frac{2}{3}$$

As $Y = \frac{2}{3}$; Verb is in agreement with the first person

Thus the singular form of verb in first person (was) has been used.

Coding for the specified work

(a)   Association of verbs with subjects:

```
str = Text1
subject = "x"
rec.MoveFirst
While Not rec.EOF
comp = " " + rec.Fields(0) + " "
comp1 = " " + rec.Fields(1) + " "
comp2 = " " + rec.Fields(3) + " "
comp3 = " " + rec.Fields(2) + " "
comp4 = " " + rec.Fields(4) + " "
f = InStr(1, str + " ", comp)
If f > 1 Then
chk = False
f1 = f
verb1 = comp
subj = Left(str, f)
obj = Right(str, (Len(str) + 2 - (f + Len(verb1))))
hv = "empty"
End If
f = InStr(1, str + " ", comp1)
If f > 1 Then
chk = False
f1 = f
verb1 = comp1
subj = Left(str, f)
obj = Right(str, (Len(str) + 2 - (f + Len(verb1))))
hv = "past"
End If
f = InStr(1, str + " ", comp2)
If f > 1 Then
chk = False
f1 = f
verb1 = comp2
subj = Left(str, f)
obj = Right(str, (Len(str) + 2 - (f + Len(verb1))))
hv = "past"
End If
f = InStr(1, str + " ", comp3)
If f > 1 Then
```

chk = False

f1 = f

verb1 = comp3

subj = Left(str, f)

obj = Right(str, (Len(str) + 2 - (f + Len(verb1))))

hv = "empty"

End If

f = InStr(1, str + " ", comp4)

If f > 1 Then

chk = False

f1 = f

verb1 = comp4

subj = Left(str, f)

obj = Right(str, (Len(str) + 2 - (f + Len(verb1))))

hv = "empty"

End If

rec.MoveNext

Wend

  (b) **The retrieval of proper verb with suitable subjects**

  If hv = "empty" Or hv = "past" Then

subject = subj

Else

s = Len(subj) - (Len(hv) + 1)

subject = Left(subj, s)

End If

If subject = " i " Or subject = " I " Then

su = Left(subject, Len(subject) - 2)

Else

su = subject

End If

  (c) **Authentication** :

 hv = "had"

sent7 = subject & " " & hv & " " & verbs & " " & obj

hv = "had been"

sent8 = subject & " " & hv & " " & verbs & " " & obj

CONCLUSION

Our attempt to denote linguistic phenomenon of English Grammr rules with the help of abstract mathematical entities and their algebraic operations based on axioms and laws of source language to the natural language domain. The mathematical formulation has enabled us to develop a computer program of the process in visual basic as front end and C++ as back end. It will help the users of English language Grammar to test their skills and knowledge of English Grammar electronically. This work is a module of our aim of developing online Grammar checker in near future.

REFERENCES

1. Casadio Claudia and Lambek Joachim (editors) (2008) Computational algebraic approach to Natural language, Polimetrica, International Scientific Publishers.

2. Chomsky N. (1957) Syntactic structures, The Hague/Paris: Mouton.

3. Harris Zellig Sabbettai (1982) A Grammar of English on Mathematical Principles, John Wiley and Sons, New York.

4. Lambek Joachim (2004) A Computational algebraic approach to English Grammar, Syntax, 7(2), 128-147, Wiley Online Library.

5. Lambek Joachim (2008) From word to sentence: A computational algebraic approach to Grammar, Polimetrica, International Scientific publishers.

6. Kumar Akshat and Shivshanker B.Nair (2007) An artificial immune system based approach for English Grammar checking, Lecture notes in Computer Science, 4628/2007,248-357, Artificial Immune system, 6th International Conference ICARIS, Springer-Verlag.

7. Pawet Scheffer and Marcin Cinciate (2010) Explicit Grammar rules and L_2 acquisition, ELT online Journal, Oxford University Press.

8. Yingxu Wang (2009) A formal syntax of natural languages and the deductive grammar, Fundamental Informaticae, 90(4), 353-368.

9. Yingxu Wang (2010) Cognitive informatics and Denotational mathematical means for Brain Informatics, Brain Informatics, Lecture notes in Computer Science, 6334/2010,2-13, Springer.