Towards an Affix Grammar for the Hungarian Language*

Ernő Farkas[†] Cornelis H. A. Koster[‡] Peter Köves[§]
Mátyás Naszódi*

Abstract

This paper is concerned with the feasibility of using Affix Grammars in describing the morphology and syntax of the Hungarian language. Two particular problem areas are studied:

1. The Hungarian morphology, with its assonance rules and highly complex congruence rules
2. The Hungarian verbal phrase, with all its concomitant problems of constituent order.

In both of these areas, we attempt to derive formal grammars rather than programming recognition automata, since the latter can be generated from the grammar by automatic means.

1 Introduction

This article is the result of a study, which took place in the spring of 1991 at the Technical University of Budapest, to investigate the applicability of AGFLs (Affix Grammars over a Finite Lattice, see [12]) to the description of the morphology and syntax of the Hungarian language.

The Hungarian language, which is highly inflected and has a relatively free word order, presents quite formidable obstacles to a syntactic description by means of a Phrase Structure Grammar.

1.1 Affix Grammars over a Finite Lattice

The Affix Grammars, which belong to the family of two-level grammars, are a formalization of the notion of augmented Context-Free grammar. The *production rules* of an Affix Grammar are Context-Free rules extended with attributes as parameters, where the *domains* of the atributes are defined by a Context-Free *meta-grammar*. In the case of

*Appeared in: A. Gyorgy (Ed.), Proceedings CIS'91 conference on intelligent Systems, John von Neumann Society for Computing Sciences, Budapest pag. 233-246, 1991

[†] Computer and Automation Institute of the Hungarian Academy of Sciences

[‡] Visiting the Technical University of Budapest on a TEMPUS-grant.

[§] Free-lance advisor

the AGFL formalism to be used in this article, the meta-grammar is particularly simple, because the domain of each attribute[1] is a choice out of a finite enumeration. For a more precise description, we refer the reader to [12]. AGFLs are equally suited for the expression of constituent order and agreement between constituents of a sentence.

Affix Grammars hve been applied in the morphosyntactic description of a number of languages:

- English (J. Aarts and N. Oostdijk [13])
- Spanish (J. Hallebeek [7, 8])
- Modern Standard Arabic (E. Ditters [4]), and
- Dutch (J. Honig, J.J. Schoorl and S.Bender [15])

It is our hope to start a similar project describing Hungarian.

In this particular paper, we investigate some aspects of the morphology and syntax of the Hungarian language with a view to their description by an AGFL. We want to highlight a number of linguistic phenomena which distinguish Hungarian from that of many other languages, and therefore may be of interest to linguists.

2 Describing Hungarian morphology

In syntactically describing a natural language, we have to choose what class of sentences we wish to describe:

- all *decypherable* sentences (necessitating a very loose description)
- all *acceptable* sentences (acceptable to whom? To an "average speaker"? A pedantic high school teacher? A poet?)
- all *correct* or *well-formed* sentences? (may lead to a circular argument)
- all *likely* sentences? (how do we measure likelyhood?)

As usual we shall take a pragmatic attitude, aiming initially at the morphosyntactic description of acceptable Hungarian sentences. We do not so much try to develop new linguistic theories as try to integrate existing theories into one consistent and relatively complete formal description. Later on, we shall use *Corpus Linguistics* methods to refine the description. In this way, existing theories can be validated or refuted, new hypotheses can be formulated and tested and a starting point can be found for more quantitative studies, relating to the syntactic likelyhood or *plausibility* of sentences.

In other words, initially we aim at an *analytical* description, which will be refined later into a more *generative* one.

[1] In an Affix Grammar the attributes are considered as meta-affixes to the nonterminal symbols of the grammar. In order to prevent confusion, we will use for that purpose in this article the term *attribute*, reserving the term *affix* for a pre-, in- or suffix according to conventional grammatical terminology.

2.1 About the Structure of Hungarian words

The Hungarian language is a typical agglutinative language. This means that several suffixes may follow the word stem and some prefixes may be used. The maximum number of suffixes in a word form is bounded by the human comprehensibility. In practice, a word stem is rarely followed by more than five suffixes but, taking into account the large number of different suffixes, a word may appear in several thousand forms in real text. For example the word *megnézethettétek* (you could make them look at it) is explained in the following way:

meg-	verbal prefix of perfection
-*néz*-	verbal root
-*et*-	factive suffix
-*het*-	suffix of ability
-*t*-	mark of past tense
-*é*-	3-rd person definite object
-*t*-	2nd person subject
-*ek*	plural of subject

As our example shows, a verbal ending expresses the number and person of the subject and the mood of the verb, but also the person of the object. In the same way, a nominal ending may be separated into possession, possessive, plural marks and case-affix. Although the functionally different marks in a nominal/verbal ending are used orthogonally, their separation is not trivial.

In Hungarian, derivative suffixes are used very productively. Contrary to the nominal and verbal endings which are enumerable, you never can predict the possible sequences of the derivative suffixes. From the point of view of morphology, a derivative suffix transforms the word from one grammatical class into another. For example, the suffix -*ít* makes a verb from an adjective (*kék*=blue, *kékít*=make/paint something blue). As the partition of the nominal/verbal ending is algoritmically too complicated – even linguists can't agree how to separate verbal endings – a more pragmatical model is used. In this model there are three (grammatical) classes of suffixes, namely: nominal endings, verbal endings and derivative suffixes. The syntax of a nominal word can easily be written in AGFL:

```
possessor:: NO, S1, S2, S3, P1, P2, P3.
number:: SING, PLUR.
possessive:: NO, SING, PLUR.
case:: SUBJ, OBJ, ..... (22 different cases)
```

These meta-rules define the domains of the attributes in the rule

```
Nominal(possessor, number, possessive, case):
   WordStem(Nominal),
   NominalEnd(possessor, number, possessive, case).
```

Introducing some more attributes we can write

```
mood:: INDICATIVE, IMPERATIVE, CONDITIONAL, PAST, INFINITIVE.
object:: NO, 1, 2, 3.
subject:: NO, S1, S2, S3, P1, P2, P3.

Verbal(mood, object, subject):
   WordStem(Verb), VerbalEnd(mood, object, subject).

WordClass1:: VERB, NOUN, ADJECTIVE, NUMBER ....
WordClass2:: VERB, NOUN, ADJECTIVE, NUMBER ....
Nominal:: NOUN, ADJECTIVE ....

WordStem(WordClass1):
   Root(WordClass1);
   WordStem(WordClass2), Derivative(Wordclass2, WordClass1).
```

The prefixes may also be handled in this way. In Hungarian there are two types of prefixes: verbal prefixes, and the prefix of the adjective for expressing the superlative. For this purpose the number of wordclass categories must be increased. Instead of `VERB` two categories should be distinguished: *VERB0* and `VERB1`, where the first one stands for verbs without verbal prefix, and the second one stands for verbs with verbal prefixes.

```
Verb::VERB0,VERB1.

WordStem(VERB1): VerbalPrefix, WordStem(VERB0); .
```

The superlative prefix can be defined in a similar way.

While in the case of nominal/verbal endings the attributes were used to mark grammatical affixes, for derivatives and word stem/word root there are grammatical attributes marked in the description.

The AGFL formalism is at first sight to be a powerful description mechanism. If all 924 (7⋆2⋆3⋆22) nominal endings and the 140 (5⋆7⋆4) verbal endings are defined, and all words and derivative suffixes are categorized, the problem of Hungarian morphology may appear to be solved.

This leaves however the problem, that not all the combinations of the affixes are legal. For example, in

```
VerbalEnd(mood,object,subject)
```

the subject may be undefined (=`NO`) only in infinitive mood. On the other hand there are affix combinations which have common representations. For example, the pairs:

```
NominalEnd(possessor,PLUR,SING,case)
NominalEnd(possessor,PLUR,PLUR,case)
```

or

```
VerbalEnd(PAST,S1,NO)
VerbalEnd(PAST,S1,3)
```

have no different representations in the Hungarian morphology. In these cases the ending-form is ambiguous.

2.2 Phonological attributes

A more serious problem comes from the fact that morphemes (word roots, derivatives, and other suffixes) may have more (1-5) different surface forms. And what is more, in most cases only one of them is legal in a given situation. Bad choice of surface-forms may cause bad generation or misunderstanding of the words. We have to introduce into the AGFL some phonological attributes to define the legal suffix forms.

This phonological classification may be factorized from the point of view of inheritance. Some phonological attributes do not change when a new suffix is added to the word (and the suffix has the same feature), but there are also phonological attributes which change from affix to affix. The second type of attributes serve as a "matching" code. Only one phonological feature has been found which is transmitted from morpheme to morpheme, namely the main vowel-harmony class. To express these features, the following meta-rules are introduced:

```
VowelClass:: FRONT, BACK.
```

and for the other phonological attributes:

```
Matching0:: M1, M2, M3.....
Matching1:: M1, M2, M3.....
```

Now all word roots, derivative suffixes and nominal/verbal endings have to be classified by these attributes. For example:

```
Nominal(possessor,number,possessive,case):
   WordStem(VowelClass, Nominal, Matching0),
     NominalEnd(Matching0,VowelClass,possessor,number,possessive,case).

WordStem(VowelClass, WordClass, Matching0):
   Root(VowelClass, WordClass, Matching0);
   WordStem(VowelClass, WordClass1, Matching1),
     Derivative(Matching1,VowelClass,Wordclass1,WordClass,Matching0).
```

2.3 Classification of the Morpheme-forms

The answer to the question how many classes there are is not trivial. It depends on the level of detail in the phonological-grammatical model that is used. There are models [5, 14] where several hundred groups are used. If such a strict model were to be used the number of rules would increase enormously. Even with a coarser phonetic classification however a good result may be reached. In particular the model accepts all wellformed words, and it finds the right interpretation. On the other hand, some bad interpretations may be generated, and misspelled word forms may be accepted as well. Now five matching classes are used, and the number of bad interpretations is not too high compared to the number of really ambiguous word forms. The bad alternative interpretations are to be eliminated at the syntactic or semantic level. Such a morphological grammar is *analytical*, rather than *generative*.

When the morphological grammar is used to generate word forms, due to the coarse classification many of them are strange, even though understandable for a native Hungarian user:

bad word form	instead of
pénzöm	*pénzem* (my money)
aludottam/alszattam/altam	*aludtam* (I slept)

2.4 Classification of the Roots and the Suffixes

Because of the problems mentioned above, an item in the lexicon is not simply the list of the word roots. All the forms coming from one word root have to be classified according to their grammatical, phonetic and matching classes. The classification of the suffixes may be solved once and for all, because of their (relatively) low number. All their possible forms can be listed together with their attributes

```
Derivative(M1, BACK      , Nominal, VERBO, M1): oz ; .....
Derivative(M1, FRONT     , Nominal, VERBO, M1): ez; öz ; .....
Derivative(M1, VowelClass, Nominal, VERBO, M2): z ; .....
```

but the classification of the word roots is a more difficult problem. One of the problems is, that there does not exist a reliable and machine readable vocabulary of Hungarian, in which all the root forms are listed and categorized. A word root may have up to 5 alternative forms, and their classification is a job for linguists. Most of the alternatives can be generated from the original vocabulary form of the word, since the root changes can be expressed by algorithmically defined transformations (shorten the last vowel, split out the last vowel, double the last consonant, eliminate the triple consonant, change the last letter *t* to *s* etc.) Such string-manipulations can hardly be expressed in an AGFL, so all root forms should be listed in the dictionary, and the generation and the classification of forms should be done outside of the AG. For example, for the verb *eszik* (eat) four root forms are to be generated:

```
Root(FRONT,VERBO,M1):
   esz ; ....      ( -em, -ik, ........)
Root(FRONT,VERBO,M3):
   ev ; ....      ( -ő, -ett, -és.....)
Root(FRONT,VERBO,M4):
   e  ; ....      ( -het, -tet, -ttem.)
Root(FRONT,VERBO,M5):
   en ; ....      ( -ni, -nök ....)
```

3 The Hungarian nounphrase

In the Hungarian language there are some components which are connected only via attributes, whereas other constructs are connected positionally. The noun phrase can be seen as a long list of components wherein each element may be empty. If the list is not empty the last element wears the case affix and the possessive affixes, independently from its word class. So we have only one conjugation and it is applicable for nouns, adjectives, numbers, pronouns etc. However the Hungarian conjugation is not too simple, we have approximately 22 case affixes and 200 case postpositions. In the literature different numbers are given about the case affixes because different linguists have different criteria which affix is a case affix. We consider an affix to be a case affix if there is verb which assigns one of its arguments though this affix. A Hungarian affix or postposition expresses the same as a German or Russian preposition and case together.

```
NounPhrase(case,numb,def):
  Reference(numb1,case,def),
    SimplePhrase,
      Plural(numb1), PossessorMark(numb), CaseTag(case).

NounPhrase(case,numb,DEFINITE):
  Possessor(pers1,numb1),
    SimplePhrase,
      PossessiveAffix(pers1,numb1,numb), CaseTag(case).
```

The noun phrase has two main parts, the first is a Reference. If the noun phrase is definite then it is a definite article preceded by an optional *ez* or *az* (this or that) which repeats the number and the case of the simple noun phrase. In the indefinite case it is an indefinite article or empty.

The other case occurs when the noun phrase has a possessor. In this case the noun phrase is always definite. The possessor and the possessive affix of the noun phrase must be in agreement. The possessive affix shows the person and the number of the possessor and the number of the possessed object.

```
SimplePhrase:
  Quantor, Selector, Quantity, Features, Name,
    Ideology, Profession, Noun.
```

Each component is optional but at least one must be there.

The quantor is a word like “all”, “each”, “certain”, etc. The selector is a serial number, or a compound verbal attribute together with some verbal arguments. The quantity is a definite or indefinite number and an optional measure. The feature is a adjective or a verbal attribute. The ideology is something like “catholic”, “communist” or “vegetarian”.

4 The verb and its arguments

English grammars usually state that the English sentence has the following structure:

```
subject verb object1 object2 prepositional objects.
```

This means that some grammatical roles are expressed by the position. This can be expressed in a CF grammar:

```
Sentence: NounPart, VerbalPart.
```

The verbal part has a different form depending on whether the verb is transitive, has an indirect object and so on.

In Hungarian the word order plays a completely different role. The sentence has the following parts:

- Topic: a reference to well-known information

- Focus: the most important new information
- Comment: additional information.

The Focus has a special stress in the accent, it is either the finite verb or a noun phrase standing before the finite verb.

Since the CF grammar is suitable for expressing the word order and not suitable for expressing other relations, we can define the Hungarian sentence in the following form:

```
Sentence: Topic, Focus, Verb, Comment.
```

or

```
Sentence: Topic, Verb, Comment.
```

if the Verb is in the Focus.

But the question arises: how can we express that grammatically the sentence is composed from the verb and its arguments and from the free adverbials. The answer is very simple: through the agreement of attributes. The Verb has a lot of features – not only the person and the number of the subject, the kind of the object, the tense, but it also determines the possible arguments. One way to express this in an AGFL is the following:

```
Sentence: Topic(cases2, number, pers, def, tense),
   Verb(cases1, number, pers, def, tense),
   Comment(cases3, number, pers, def, tense),
   whereCompatible(cases1, cases2, cases3).
```

This definition can be interpreted as follows: In the head of the sentence there are some components forming the Topic. If the subject is in the Topic it must agree with the Verb in person and number. If the subject is in the Topic it must agree with the Verb in definiteness and whether it is the second person or not. If a time adverbial is in the Topic it must agree with the tense of the Verb. Similar agreements hold for the Comment. The `cases2` and `cases3` attributes determine the component set. The last element `whereCompatible` is a check. It checks whether the components which are given in the Topic and in the Comment are those which are required as the argument of the Verb or a free adverbial.

This leaves the problem that the cases1, cases2, cases3 are set or list attributes. A verb usually has 0 to 5 arguments, so we can enumerate all the case frames and all subsets. But this is only hypothetical possibility since in practice it is an extremely big number.

A possible solution might be to introduce set valued attributes with the associated set operations over the attributes. Then we could write something like:

```
Topic(cases+case,...): Component(case,...), Topic(cases,...).

Topic(0,...): .
```

where the `0` denotes the empty set and the `+` denotes that an element is joined to the set.

It is not the intention to introduce such computing power into the AGFL formalism, which describes purely relations between components. In the next section, another solution is investigated.

5 The problem of free word order

Often it is stated about some language (especially by non-linguists) that “the word order is free”. Certainly there may be a remarkable degree of freedom in the relative order of constituents in a sentence, but there are hardly any languages where a shuffle of the words in any sentence leaves its meaning invariant.

5.1 Word order in Hungarian

In [9], S.Karoly states

> The word-order structure of the Hungarian sentence is very flexible in adapting itself to the requirements of communication: it has nearly maximally developed means in distinguishing new or known, emphatic or non-emphatic part or parts by word order or stress in speech. This results in the fact that basically freedom is characteristic of the Hungarian word order; we can say that the Hungarian word order is free.

The author then presents 25 different word order permutations for the following relatively simple sentence:

> a fiam elküldte a könyvet a barátjának
> “my son sent the book to his friend”

Considering this example as a 7-word sentence one would expect 7! different orders if word order is indeed free. Inspection of his examples show rather that the sentence consists of 4 components (subject, verb, object and prepositional object) which can be given in any order – thus, one would expect 24 permutations.

1. a fiam elküldte a könyvet a barátjának
2. a fiam elküldte a barátjának a könyvet
3. a fiam a könyvet elküldte a barátjának
4. a fiam a könyvet a barátjának elküldte
5. a fiam a barátjának elküldte a könyvet
6. a fiam a barátjának a könyvet elküldte
7. elküldte a fiam a könyvet a barátjának
8. elküldte a fiam a barátjának a könyvet
9. elküldte a könyvet a fiam a barátjának
10. elküldte a könyvet a barátjának a fiam
11. elküldte a barátjának a fiam a könyvet
12. elküldte a barátjának a könyvet a fiam
13. a könyvet a fiam elküldte a barátjának
14. a könyvet a fiam a barátjának elküldte
15. a könyvet elküldte a fiam a barátjának
16. a könyvet elküldte a barátjának a fiam
17. a könyvet a barátjának a fiam elküldte
18. a könyvet a barátjának elküldte a fiam
19. a barátjának a fiam elküldte a könyvet
20. a barátjának a fiam a könyvet elküldte
21. a barátjának elküldte a fiam a könyvet
22. a barátjának elküldte a könyvet a fiam
23. a barátjának a könyvet a fiam elküldte
24. a barátjának a könyvet elküldte a fiam

To a native speaker of Hungarian, all these variations are indeed acceptable, in the sense that they may occur as sentences in some suitable context, but that does not mean they are all syntactically and semantically equivalent!

To begin with, those starting with the verbform are questions, rather than statements, even though in verse or highly poetic prose they might occur as statements. Secondly, the initial position in the sentence shows the *topic* of the sentence – the various word orders differ in topicalization. It is a moot point whether this is a syntactic or a semantical difference. Lastly, some sentences look more likely to the native speaker, whereas others make in this form a contorted impression. There are important stylistic differences between the different word orders, and style is to a large extent expressed by syntactic means (see e.g. [3]).

Curiously, Karoly gives only 7 of the above 24 permutations. He gives 18 more, but these are not pure permutations of the original sentence. The reason is, that he takes into account not only word order but also *emphasis*, usually indicated by the *stress*, a phonological attribute. Although the stress is invisible at the morphological level, emphasis may at least partially find syntactic expression: in this sentence the verbal prefix *el* may be turned into a postposition to indicate that the preceding complement is emphasised. This gives rise to 18 more variations.

25. a fiam küldte el a könyvet a barátjának
26. a fiam küldte el a barátjának a könyvet
27. a fiam a könyvet küldte el a barátjának
28. a fiam a könyvet a barátjának küldte el
29. a fiam a barátjának küldte el a könyvet
30. a fiam a barátjának a könyvet küldte el
31. a könyvet a fiam küldte el a barátjának
32. a könyvet a fiam a barátjának küldte el
33. a könyvet küldte el a fiam a barátjának
34. a könyvet küldte el a barátjának a fiam
35. a könyvet a barátjának a fiam küldte el
36. a könyvet a barátjának küldte el a fiam
37. a barátjának a fiam küldte el a könyvet
38. a barátjának a fiam a könyvet küldte el
39. a barátjának küldte el a fiam a könyvet
40. a barátjának küldte el a könyvet a fiam
41. a barátjának a könyvet a fiam küldte el
42. a barátjának a könyvet küldte el a fiam

Still, this example is an impressive demonstration of the liberal order of constructs in Hungarian.

5.2 Describing permutations

We consider a model language, in which a verb $\mathtt{v}$ takes some number of arguments $\mathtt{a}_1, \ldots$ $\mathtt{a}_n$, where both the number n of arguments and the syntactic category of each argument depend on the *class* of the verb. As an example, some of the verb-types that may be distinguished in English are indicated in the following table:

Verb	object	auxiliary	prep.phrase
to exist	no	no	no
to live	maybe	no	no
to see	yes	maybe	no
to owe	yes	no	TO
to hit	yes	maybe	no
to abide	no	no	WITH

allowing such sentences as

I exist
I live (my own life)
I see you (with my own eyes)
I owe my life to his presence of mind
I hit him (with a blunt object)
abide with me.

In our model language, in distinction to English, the arguments may surround the verb (precede and follow it) in any order. The structure of its verbal phrase can therefore be characterized by the following syntax:

```
verbal phrase: comps, verb, comps.

comps: comp, comps; .

comp: subject; object; auxiliary; prep phrase; etc.
```

This is an unnecessarily ambiguous syntax. Rewriting the first rule to

```
verbal phrase: comp, verbal phrase; verb, comps.
```

we obtain a left-factored form, which can easily be recognized from left to right in linear time.

In recognizing a sentence according to this grammar, however, we attribute no structure at all to it, apart from finding the boundaries between its constituents. In particular, the fact that the verb governs the number and categories of the arguments is not described in the grammar, nor is it checked during syntax analysis. Checking the government is delegated to a later, semantical phase, even though it is a syntactic matter. This is highly unsatisfactory.

Another approach is to take careful stock of all verb-classes that occur (say m) and then write a grammar according to the pattern

```
verbal sentence:
   pattern of class1;
   pattern of class2;
   ...
   pattern of class m.

pattern of class 1:
   subject, verb;
   verb, subject.
```

and so on, enumerating each of the permutations for each of the verb classes independently. Again, a left-factorization may be helpful in recognizing this language more efficiently; let us assume that this is done automatically. The main objection against this solution is that it is plain boring: the enumeration of all possible permutations takes a number of lines that is exponential in the number of components, and even if we get the enumerations right this is a very unsatisfactory way to present a linguistic theory of verbal sentence structure.

5.3 The GPSG solution

In the book on Generalized Phrase Structure Grammar by Gazdar, Klein, Pullum and Sag [6] a distinction is made between immediate *dominance*, the relation between a mother (i.e. left hand side) category and its daughter (i.e. right hand side) categories, and linear *precedence*, the relations between daughters of one same mother. In GPSG, the notation

```
A --> B, C, D
```

specifies immediate dominance: an A consists of a B, a C and a D, *in any order.* Any restrictions on the order can be specified separately, e.g.

$$B \prec D$$

Which constrains the B to always precede the D, effectively addmitting the orders:

```
B C D
C B D
B D C
```

and no others[2]. In the extreme case where the precedence admits only one order (which is the usual case in CF grammar) the Immediate Dominance rule has to be accompanied by a Linear Precedence rule over the same elements:
$B \prec C \prec D$
so we end up writing twice as much as in a CFG.

The claim made about the ID/LP mechanism that it captures generalizations not expressed in a CFG is a moot point. The examples in the book all allow a CF treatment that is just as clear as the GPSG description. The Exhaustive Constant Partial Ordering property that GPSG grammars have to satisfy is difficult to fathom, a piece of learned Mumbo-Jumbo rather than a law of Nature.

5.4 An extension to AGFL

In trying to apply ID/LP to Hungarian, there is clear evidence of Immediate Dominance but none of (constant) Linear Precedence. It seems wise to adopt a specific notation for the case that the members of an alternative can occur in any order (corresponding to an unrestricted immediate dominance), so we let

```
A : B | C | D.
```

stand for the case where all of B, C and D have to be present as immediate constituents of A, in some order whatsoever.

The operator `|` (*permutation choice*) has a higher priority than the comma (*concatenation*). Considered as operators on languages, these operators possess the following properties (here ϵ stands for the language consisting of only the empty string:

[2] Gazdar et al. strongly suggest that conventional CF grammar is not capable of expressing such ordering restraints, although this is actually quite simple, albeit rather boring.

```
a : B, cd; C, bd. cd: C, D; D, C. bd: B, D.
```

$$\begin{array}{rcll}
a, \epsilon & = & a & (\mathit{empty\ 1}) \\
\epsilon, a & = & a & (\mathit{empty\ 2}) \\
a|\epsilon & = & [a] & (\mathit{option}) \\
a|b & = & b|a & (\mathit{commutativity}) \\
a,(b,c) & = & (a,b),c & \\
a|(b|c) & = & (a|b)|c & \\
a,(b|c) & = & (a,b)|(a,c) & \\
(a|b),c & = & (a,c)|(b,c) &
\end{array}$$

By means of this extension to AGFL we intend to deal with the word order problems that were encountered. It allows us to avoid tedious enumerations of all possible orders of components of a pattern. Of course, it is not possible to avoid by this device the enumeration of all verbal class patterns in Hungarian (this is after all a distinctive aspect of Hungarian syntax) but each of them can be defined by one succinct rule.

6 Conclusion

We have performed only a small-scale experiment, so the results must be seen as very tentative.

We have tried the AGFL-notation on a small subset of the Hungarian syntax, with a limited vocabulary for which we can enumerate all the case frames. The generative result was very interesting, we find that many good and some bad statements were generated in our model. We recognized that some verbs and verb forms must have or must not have an accentual stress, so that the verb has an other attribute, the aspect, which determines the sentence patterns in which the verb can occur.

An Affix Grammar is convenient for handling grammatical affixes on the syntax level. It is useful for describing the structure of a word (morphosyntax), including some phonetic attributes of the morphemes, however the number of rules may be very high. The most difficult problem (which is not solved for Hungarian) is the classification of the Hungarian words from the point of view of phonetics and alternative root forms. This problem is not a matter of choice of notational formalism.

The experimental morphological analyzer written in AGFL is about 40 KB long, excluding the lexical part. The quality of the analyzer is acceptable, but as a generator it needs some more refinements, since it presently generates more “foreign” word forms than right solutions.

In the application of AGFLs to Hungarian, a non-Indoeuropean language, new problems were to be expected, which might demand the development of novel solutions. In particular, the relatively free word order of Hungarian is hard to express in a traditional phrase structure grammar. We have seen that an extension to the formalism is indeed required to deal with this phenomenon in a satisfactory fashion.

References

[1] Jan AARTS and Theo van den HEUVEL, *Computational tools for the syntactic analysis of corpora.* In: Linguistics 23, 303-335, 1985.

[2] Loránd BENKŐ and Samu IMRE (Eds.), *The Hungarian Language*, Akadémiai Kiadó, Budapest, 1972.

[3] Chrysanne DIMARCO, *Computational Stylistics for Natural Language Translation*, Technical Report CSRI-239, University of Toronto, May 1990.

[4] Everard DITTERS, *A Formal Grammar for Automatic Syntactic Analysis and other Applications*, In: *Proceedings of the Regional Conference on Informatics and Arabization*, IRSIT, Tunis, Vol.1, 128-45, 1988.

[5] Lászlo ELEKFI, *Közyelvi kiejtésünk és az Ertelmezö Szótár*. Szótártani tanulmányok, Budapest, 1966.

[6] Gerald GAZDAR, Ewan KLEIN, Geoffrey PULLUM and Ivan SAG, *Generalized Phrase Structure Grammar.* Harvard University Press, Cambridge Mass, 1985.

[7] Jos HALLEBEEK, *Hacia un sistema de análisis sintáctico automatizado: el proyecto ASATE.* In: Martin Vide, C. (Ed), *Actas del II Congreso de Lenguajes Naturales y Lenguajes Formales.* Universidad de Barcelona, 545-558, 1987.

[8] Jos HALLEBEEK, *Een grammatica voor automatische analyse van het Spaans.* Diss. University of Nijmegen, 1990. In Dutch, French translation to appear.

[9] Sándor KÁROLY, *The Grammatical System of Hungarian*, In [2].

[10] Katalin E. KISS, *Configurationality in Hungarian.* Reidel, Dordrecht and Akadémiai Kiado, Budapest, 1987.

[11] András KOMLÓSY, *Régensek és Vonzatok.* Kandidatusi disszertácio, Budapest, 1991.

[12] Cornelis H.A. KOSTER, *Affix Grammars for Natural Languages*, In: Henk Alblas and Bořivoy Melichar (Eds.), Proceedings Summer School on Attribute Grammars and Applications, to appear in Springer Lecture Notes, 1991.

[13] Nelleke OOSTDIJK, *An Extended Affix Grammar for the English Noun Phrase.* In: Jan Aarts and Wim Meijs (eds), *Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research*, Amsterdam: Rodopi, 1984.

[14] Ferenc PAPP, *A magyar fönévragozás három modelije.* CMagyar Nyelv, 1966.

[15] Jan J. SCHOORL and Simon BELDER, *Computational Linguistics at Delft, a Status Report*, Report WTM/TT 90-09, Delft University of Technology, 1990.