PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The version of the following full text has not yet been defined or was untraceable and may differ from the publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/18691

Please be advised that this information was generated on 2019-04-22 and may be subject to change.
NLCA - A truly relational model of language

V. Kamphuis, J.J. Sarbo

Computing Science Institute/

CSI-R9815 May 1998
NLCA — A truly relational model of language

V. Kamphuis and J.J. Sarbo
University of Nijmegen, The Netherlands

Abstract

The primary goal of Natural Language Concept Analysis (NLCA) is to find a means of handling flexibility in natural language. Contrary to traditional approaches using the part-whole paradigm (e.g. phrase structure grammar), NLCA identifies a relational basis underlying hierarchical structure. This basis derives from the interaction of lexical items due to their inherent combinatorial properties, and constitutes a set of three relation schemes: major predication, minor predication, and qualification. A comparison between NLCA and dependency-based description reveals their common basis in word-based modelling, but shows fundamental differences in their foundation.

1 Introduction

Many mainstream approaches to natural language modelling use some form of phrase structure in the specification. Examples are HPSG [PS94] and GB [Hae91]. Phrase structure is traditionally considered an important theoretical notion in linguistics. However, the underlying reasons for its strong presence have not really been clarified. Current research [DS98] reveals that the phenomenon 'phrase structure' is by no means accidental, as in the Peircean semeiotic view, such structure is a realization of signs that show some form of completeness.

In spite of its strong theoretical status, practical applications based on phrase structure description are not without problems. The main reason for this is the high flexibility of natural language. In performance data (i.e. actual language use), many disruptions of, and variations on standard phrase structure patterns occur. The problematic nature of such linguistic phenomena can be classified in terms of the typical cases of (a) Flexibility, (b) Structural variation, and (c) Linguistic vagueness.
Flexibility problems refer to the interruption of standard phrase structure patterns, as is the case with insertion, discontinuity, non-phrase structure patterns occurring in the context of e.g. coordination and comparison, and incompleteness. Structural variation covers cases of variation in order of constituents, and also variation in instantiations of a pattern. Finally, linguistic vagueness relates to unclear constituency relations and unclear lexical classification.

In grammar-based specifications, handling phenomena of this kind usually involves a repetitive process of extending and adjusting the grammar. Experience shows that such phenomena occur frequently and that their instances show large unpredictability. This means that ad hoc adjustment of the grammar is almost inevitable. This may lead to serious maintainability problems, eventually. There is also a theoretical issue involved: how is one to know that the grammar has complete coverage of the language, that is, when is the grammar complete? The possibility exists that no more than ‘accidental’ coverage is accomplished.

Because of the problematic nature of hierarchical structure, there is a continuing search for alternative methods. Among these alternatives, dependency-based models choose to abandon phrase structure altogether. This has obvious advantages, but does not solve all problems. For example, in the description of what constitutes a noun phrase, a choice has to be made between the determiner or the noun as the head. Such a choice cannot be made in a principled way, as neither element is obligatory in all cases.

In NLCA we take the view that hierarchical levels are created by the interaction of different relations between elements, as opposed to a single type of relation (e.g. constituency or dependency) only. Because of this, it is not the presence of a particular item that hierarchical structure is based on. In general, we consider hierarchical structure in language to be the result of a dynamic process in which the interaction of different relations reaches some form of completeness.

Our model capitalises on the property that the information carriers, the lexical items, are ‘willing’ to combine. These combinatorial properties are determined by inherent characteristics of lexical items. Hierarchical structure follows naturally from the interaction of these properties, while leaving room for variation and flexibility in structural patterning.
2 A description of the model

NLCA [KS98] defines three relation schemes underlying structure in language: major predication (MP), minor predication (mp), and qualification (Q). The three relation schemes can be recursively applied, and their sum uniquely characterises the input. The relation schemes MP and mp are referred to uniformly as predication. Predication is a pair \((p; a_1, \ldots, a_n)\), for \(n > 0\), where \(a_1, \ldots, a_n\) function as arguments to the predicate \(p\).

The different relation schemes reflect certain conceptual distinctions that may be expressed by means of language. The symmetrical instantiation of predication involves the distinction between an action/state and its participants: each requires the presence of the other. The asymmetrical variant distinguishes between an action/state or participant on the one hand, and its properties on the other (modification): the predicate requires the argument of which it is predicated, but the reverse does not hold. We call predicates of the first type major predicates, and predicates of the second type minor predicates. There are various distinguishing factors between major and minor predicates. In English, major predicates (usually) relate to the noun-verb division; minor predicates do not. Major predicates are typically realized by verbs; minor predicates by adjectives and adverbs. There is never more than one major predicate associated with an argument; there may be several minor predicates related to the same argument. (This reflects the possibility of having zero or more modifiers of an action or participant.)

The third type of relation, qualification, distinguishes between the core content of a linguistic expression and some qualification of it. At the level of an action and its participants, for example, this qualification may relate to referential status of NPs (e.g. definite vs. indefinite), or to tense and aspect information expressed by the verb. Intensifying adverbs (e.g. very, extremely, deeply) and comparative adverbs (e.g. more and most) also belong to the class of qualifiers.

The difference between a minor predicate and a qualifier is that the latter does not introduce a meaning that is independent of the element it qualifies. By contrast, a minor predicate has some aspect of meaning that is independent of the element it combines with.\(^1\) Furthermore, there can be several modifiers associated with an argument or predicate; typically, however, there will only be a single (possibly composite) qualifier. In the case

\(^1\)This is illustrated by the fact that minor predicates can be used in different contexts. For example, a prepositional phrase can modify an argument (e.g. noun) but also a predicate (e.g. verb). An adjective phrase can be used as a modifier of a noun, but also in the complementation of a verb.
of referential information, for instance, the qualifier situates the argument or predicate in its referential context of which there will only be one. In some cases different aspects of the qualifier can be expressed separately (such as tense and aspect); in that case these different aspects must be unifiable but there cannot be more than a single qualifier relating to the same domain.

Because the qualifier has no meaning that is independent of the element it qualifies, the presence of a qualifier also signals the core associated with its type. In case the qualifier precedes the core, this is modelled in NLCA by introducing a placeholder, called a Proto-item, for the core. When the core is realized, it replaces the Proto-item. Only qualifiers may introduce Proto-items (based on their special relation to the core).

The three relation schemes are formalised as follows. A qualification \((Q)\) consists of a qualifier and a core. The qualifier has no information content independent of the core; it makes the core more specific. The core fulfills the combinatorial need of the qualifier. A minor predication \((mp)\) consists of a (minor) predicate and an argument. The predicate has information content independent of its argument and adds new, factual information to it. The relation between minor predicate and argument is asymmetric: the predicate needs its argument, but not the other way round. A major predication \((MP)\) consists of a predicate and its argument(s). Both have information content, and the relation between predicate and argument(s) is symmetric (each requires the presence of the other). The predicate introduces an argument structure, and incorporates its arguments into a single relation.

An instance of a relation scheme is called a relation. The \(Q\)-, \(mp\)- and \(MP\)-relations can be realized in language on different levels, e.g. on the morphological level, on the level of syntax, or on the more abstract level of features. We will come back to this in Sect. 6 below.

**Example 1** *The young man ate some plums.*

There are four \(Q\)-relations: the-man, PAST-eat, PLURAL-plum, and some-plums. The adjective young is in \(mp\)-relation with man. There is a \(MP\)-relation between the verb eat and its arguments, the young man and some plums.

As the example makes clear, lexical items may participate in several relations at the same time. The potential of relations that a lexical item may participate in is considered to be a fundamental property of that item, and is represented for each lexical item in terms of so-called argument positions associated with it. In accordance with the number of relations, there are three different types of argument position. The internal argument positions represent information about the item itself: there is one for the \(Q\)- and
one for the *mp*-relations, denoted by \( \text{int}(q) \) and \( \text{int}(m) \), respectively. The external argument positions \( \text{ext} \) of a lexical item represent its combinatorial demands. In the case of verbs there are as many of such positions as the number of obligatory arguments. The argument positions can be labelled (e.g. in the case of verbs, AGENT); the labelling is defined by the lexicon. In the graphical representation, internal and external argument positions of lexical items are given as buckets on the left and right hand side, as follows: \( \text{int}(m), \text{int}(q), \text{lexical unit \_ext} \). An empty bucket is omitted.

### 3 Evaluation of relations

In NLCA the input is analysed from left to right, and the relations are evaluated incrementally. A relation is evaluated when qualifier and core, or predicate and argument(s) bind to each other. The evaluation, which can be initiated by either participant in the relation, is greedy, meaning that lexical items relate with the nearest surrounding candidates available (greedy binding). Availability is restricted by the demand that only visible items can bind to each other. The visibility structure and any change to it, is due to the relations: each introduces a new visibility range for itself. In English, *mp*-relations do not change visibility; this coincides with the optionality of modifiers.

A visibility range is terminated by **closing** (and by encountering end-of-sentence). This operation applied to a combination of lexical items can yield a single new item, called a *lexical unit*.\(^2\) (Technically, a new lexical unit is only generated in the case of potential ambiguity on the level of representation. We will use this feature in the examples.) The lexical items involved in closing are no longer accessible for other relations. Closing is triggered when a lexical unit is encountered that cannot form a relation with visible units in any way.

We represent relations by a **Relation Matrix** (RM). There is a row allocated for each noun (called an object), and a column for each article, preposition, adjective, adverb and verb (called an attribute). Furthermore a column is allocated for each external argument position of a verb. For Proto-items a row or a column is allocated (referred to as Proto-object and Proto-attribute), depending on the qualifier introducing it. The RM forms...

---

\(^2\)N.B. a lexical item is a lexical unit; the representation and the principle described above extend naturally from lexical items to lexical units. The properties and the bindings of the lexical unit yielded by closing are derived from the constituent relations involved. This topic is not further elaborated on here, for reasons of space.
the basis for a representation reflecting conceptual structure in terms of Conceptual Graphs and Concept Lattices [Sow84], [KS98].

Both predicate–argument and qualifier–core relations are based on functor–argument relations. These are represented as pointers between lexical units, or in the case of morphological or feature realization, possibly as constants. The source of a pointer is a lexical unit; the destination is an argument position of the related item. Formally, the pointers between rows and columns are stored in the cells of the RM; in the examples however, they are graphically represented. Besides the pointers, a cell contains a ‘+’ sign if the destination of a pointer stored in the cell is an external argument position. These signs will be used as markers of the emerging phrasal structure. We say the input string is well-formed if the combinatorial need of each lexical unit is satisfied, meaning that the external argument positions of all items are filled. A clause is a well-formed string.

4 An example

We show the stepwise development of the analysis of a sample sentence. The information about lexical units is contained in the lexicon, such as type, number and location (pre- or post-) of argument(s) etc. In the example lexical information is specified on-the-fly.

Example 2 The young man from Rome ate some plums.

the generates a column and introduces a new visibility range. As a qualifier, it functions as the internal argument of its object. The qualifier precedes its core, therefore it creates a Proto-object and points at its qualifying internal argument position.

- the → Proto-object_int(q)

young generates a column. Its internal argument positions are not filled. As an adjective, its external argument position needs to be filled with a nominal element. There is a Proto-object present, hence there is a pointer from the Proto-object to the external argument position of the adjective (greedy binding); this results in a ‘+’ in the RM under ‘young’. The attribute itself points at the modifying internal argument position_int(m) of the Proto-object. Note that this leads to a chain of pointers from ‘the’ via the Proto-object to an external argument that has been filled; such a chain gives rise to inheritance of bindings to all units involved in the chain. Therefore, there will also be a ‘+’ under ‘the’ and a pointer from
the Proto-object to ‘the_ext’. This, in fact, creates the equivalent of a
nominal adjective phrase with an implied head (the Proto-object).

- Proto-object → young_ext
- young → Proto-object_int(m)
- Proto-object → the_ext
- ‘+’ in RM in cell Proto-object/young
- ‘+’ in RM in cell Proto-object/the

There is an important difference here between the role and treatment of
the article and the adjective. Note that the nominal adjective phrase would
not be created without the article: being a qualifier, it is the article that
supplies the Proto-object that functions in the nominal adjective phrase.
The adjective belongs to the class of modifiers that are involved in the
relation of minor predication. For this reason, they can be said to have
an implicit object required to fill their external argument position. The
Proto-object generated by the qualifier can fill this role.

man replaces the Proto-object. The noun inherits the bindings of the
Proto-object. There is still a phrase, but now it is a full noun phrase
rather than a nominal adjective phrase. Since there is not yet a pointer to
the external argument position of the noun, we still do not have a clause,
only a phrase.

- ‘man’ replaces Proto-object

from A preposition, generates a column. Together with its obligatory com­
plement, a preposition creates a phrase that enters into the mp-relation
that it initiates. This is modelled as follows: ‘from’ fills man_int(m), but
‘man’ does not fill from_ext. This position is filled by the prepositional
complement. With respect to the latter, the preposition establishes a MP-
relation (and a new visibility range): the complement is not available for
entering into another MP-relation, e.g. with a verb.

In sum, prepositions invoke a mp- and a MP-relation. Greedy binding of
the mp-relation with a preceding visible unit takes precedence over binding
of the MP-relation. The double function of the preposition is defined in
the lexicon, and modelled in the RM by means of the pointers associated
with it.

- from → man_int(m)

Rome A proper noun, generates a row. It has unique reference (unless
otherwise specified; see the discussion in sect. 6 below). This information
can be taken as part of the qualifying internal argument \(_{\text{int}(q)}\). Its external argument position is filled by the preposition, modelling the fact that a prepositional complement is no longer available for participation in any other MP-relation.

- Rome $\rightarrow$ from\_ext
- from $\rightarrow$ Rome\_ext
- ‘+’ in RM in cell from/Rome

**ate** generates a column and introduces a new visibility range. This triggers closing of the PP. The verb’s qualifying internal argument position is filled by the feature PAST; its external arguments are AGENT and THEME.\(^3\) Since ‘eat’ is a major predicate, it fills the external argument position of the object ‘man’, and ‘man’ points to the AGENT role. As a result, there is a ‘+’ in the Relation Matrix in cells man/AGENT and man/eat. However, since only one of the external argument positions of the transitive verb is filled, the clause is not yet complete.

- eat $\rightarrow$ man\_ext
- man $\rightarrow$ eat\_ext (THEME)
- ‘+’ in RM in cell man/AGENT
- ‘+’ in RM in cell man/eat

**some** A quantifying pronoun which may function as a determiner or as an independent pronoun. We can make a unified account if we treat it as a qualifier that, like the article, introduces a Proto-object; however, unlike with articles the Proto-object now also points to the external argument position of the qualifier. (This explains the possibility of e.g. *He ate some*, which, indeed, is complete but has an implicit object.) As a result, there is a ‘+’ in the Relation Matrix in cell Proto-object/some. The Proto-object also realizes the external argument THEME, causing a ‘+’ to be placed in the appropriate cell of the Relation Matrix. (We note that Proto-objects are not able to fill the external argument position of major predicates when the external argument of the qualifier that created them is not filled; cf. the ungrammaticality of *The walks*).

- some $\rightarrow$ Proto-object\_int(q)
- Proto-object $\rightarrow$ some\_ext
- Proto-object $\rightarrow$ eat\_ext (THEME)
- eat $\rightarrow$ Proto-object\_ext

\(^3\)Note that the intransitive variant of *eat* would not yield a successful analysis due to violation of the well-formedness condition: plum\_ext cannot be filled.
• ‘+’ in RM in cell Proto-object/some
• ‘+’ in RM in cell Proto-object/THEME
• ‘+’ in RM in cell Proto-object/eat

plums replaces the Proto-object. ‘s’ can be regarded as part of the qualifying internal argument. Note that this does not conflict with the fact that ‘some’ is also a qualifier: they are unifiable within the same domain (both can signify plural; together they are plural indefinite).

• ‘plums’ replaces Proto-object

End-of-sentence is reached; the noun phrase, and then the clause, are closed. The Relation Matrix for this sentence is displayed in Fig. 1.

Figure 1: The young man from Rome ate some plums

This treatment of quantifying pronouns has two important advantages. First, it does not require ambiguous lexical entries. The same can be said for demonstrative pronouns, numerals and other function words that are ambiguous between independent and adjectival use. Second, the use of Proto-objects makes it unnecessary to have a rule defining noun phrase heads as realized either by nouns, or by numerals, quantifying pronouns, demonstrative pronouns etc. In fact, this also applies to nominal adjective phrases: there is no need to define adjectives as possible realizations of noun phrase heads. The nominal adjective phrase follows naturally from the presence of the article (creating the Proto-object) and the adjective (combining with the Proto-object). This completes the Q-relation, making the Proto-object accessible for participating in MP-relations. Furthermore, this approach also accounts for the potential structural ambiguity of a quantifying pronoun or a nominal adjective phrase followed by a plural noun phrase, as in apposition. (Example: ‘On Monday she got a big bunch of flowers. The white, lilies, wilted after a mere few days.’)
Going through the sentence from left to right, we see the following structure emerge:

- At word ‘young’ we obtain the nominal adjective phrase (+1 and, through inheritance, +2);
- At word ‘man’ we obtain the noun phrase (‘man’ replaces Proto-object);
- At word ‘Rome’ we obtain the prepositional phrase (+3);
- At word ‘ate’ we have closing of the prepositional phrase which is incorporated in the preceding noun phrase as minor predicate;
- At word ‘some’ we obtain the clause with an independent pronoun (+6, +7 and +8);
- At word ‘plums’ we obtain the clause with ‘some’ as determiner (‘plums’ replaces Proto-object).

5 Flexibility and structural variation

In the previous sections we have discussed a model which attempts to explain the underlying nature of hierarchical structure in language. Our aim in doing so was to find a method of description which is inherently more flexible than one that takes hierarchical structure (in particular, phrase structure) as given. This is needed especially in order to account for non-phrase structure configurations that occur in natural language. In this section we shall illustrate the greater flexibility of our approach on the basis of an example of discontinuity (Ex. 3) and of coordination (Ex. 4).

**Example 3** A man entered who was covered with mud.

This example is problematic for a phrase structure-based account, because the sequence Article–Noun–RelClause is not continuous, but interrupted by the verb phrase which functions at a higher level. The description of such configurations usually requires either movement operations or some other mechanism that relates the relative clause to the hierarchical position immediately following the head of the noun phrase. In dependency-based description, discontinuity is problematic as long as the traditional Adjacency-constraint applies; under the usual assumptions regarding heads and dependents, this constraint in fact makes a dependency grammar more or less equivalent to a context-free phrase structure grammar [Fra96], [Hud96].
NLCA due to its relational basis can handle discontinuous modification without problems, as there is no need to refer to a position following the noun, only to visible candidates for entering into relations. The relative pronoun who can be treated in much the same way as the preposition discussed before. It invokes a mp-relation with man, which can only be realized provided that it participates in a MP-relation with its external argument. When the resulting relative clause is closed, the newly created lexical unit completes the mp-relation introduced by who.\(^4\)

In the example, upon finding the word who we have just completed the MP-relation between a man and entered. The item who enters into a mp-relation with man. The finite verb phrase was covered fills who.ext and at the same time provides an argument role for who. When the clause who was covered with mud is closed, the resulting new lexical unit completes the mp-relation invoked by who. This involves inheritance of the binding with man.

Our second illustration concerns coordination.\(^5\) This phenomenon often gives rise to disruption of regular phrase structure patterns, for example in cases of conjunction reduction, gapping, or other cases of non-constituent coordination. An example is the following sentence:

**Example 4** The young man from Rome ate some pancakes yesterday, and a steak today.

This example is problematic for a phrase structure-based account, because the substrings some pancakes yesterday and a steak today do not form a single unit at any level of hierarchical structure; rather, they contain two independent constituents, one of which functions as an adverbial at clause level, and the other as direct object to the verb. In standard phrase structure, such coordination cannot be described, as there is no single unit of analysis that the rule describing the coordinate structure can refer to. NLCA accounts for such examples as follows.

---

\(^4\) With independent relative pronouns, e.g., whoever, the unit that results after closing participates in a MP-relation. This distinction is reflected by classifying independent relative pronouns as participating in two MP-relations, rather than a mp- and a MP-relation.

\(^5\) Interestingly, in dependency literature coordination is sometimes mentioned as the one example for which the constituency relation might be necessary; e.g. [Hud95]. This reveals a conflict between the need to refer to larger units of analysis on the one hand, and the rigidity imposed by such larger units on the other. It is precisely this conflict that has given rise to NLCA’s quest for a more abstract principle identifying larger units. See also the discussion in Sect. 6.
In general, coordination refers to the *relations* preceding the coordinator. The coordinator separates two visibility ranges, one on the left and one on the right hand side. In the current example, upon reaching the coordinator, we have completed a MP-relation, i.e. the external arguments have been found. The effect of the coordinator *and* is that the external argument positions (in this case, of the major predicate) can be re-used. In essence, coordination connects two compatible lexical units in the left and right conjunct, and relates them to the context of the coordinate structure as a whole [Kam97]. Briefly, two units are *compatible* when they may participate in the same relation scheme, in the same role (i.e. qualifier or core; predicate or argument), and relate to the same type of lexical unit. For the individual conjuncts in the context the well-formedness condition applies.

6 Dependency and NLCA

Having a strong lexical basis, NLCA has a lot in common with dependency-based models such as Word Grammar [Hud84]. At the same time, however, there is a fundamental difference between them, as there is between NLCA and phrase structure-based approaches. This difference can be illustrated in the following manner.

In the dependency or Word Grammar approach, a sentence can be viewed as a building that has one primary foundation: the verb. Built on top of this is the next layer, consisting of the dependents of this verb. Again, the layer built on top of this consists of the dependents of these elements, and so on until every element has been described. Ultimately, then, there is one element that carries all the others in a transitive relation. In such an approach, “[a] phrase is just a group of words which all depend, directly or indirectly, on one single word, so adding phrase-units to the analysis would add no extra information whatever” [Hud95].

In NLCA, a different picture emerges. There is no unique element that can be identified as the foundation onto which all other layers are added; indeed, there are no individual building blocks at all. Rather, the sentence is built on the basis of a web of interweaving relationships motivated by inherent properties of lexical units. A larger unit is formed when these relationships have been completed.

This different view has fundamental consequences for the linguistic analysis NLCA provides. For instance, let us consider the analysis of the noun phrase. In a dependency approach, a choice has to be made between the determiner or the noun as the dependent element (much in the same way
As phrase structure-based analysis has to choose between the Determiner Phrase (DP) with the determiner as head, and the Noun Phrase. Word Grammar chooses the former, on the basis that “most determiners can be used without a following common noun”, which “suggests that the determiner is the obligatory part of a phrase consisting of a determiner plus a common noun, so it should be the determiner, rather than the common noun, that carries the external relations” [Hud95]. However, the above claim regarding the optionality of the following noun does not hold when the determiner function is realized by an indefinite article, or is not realized at all, nor can the determiner be said to be obligatory in all cases. In fact, as the following examples show, there is not one single element that can be identified as obligatory in noun phrases:

the man
men
bold men
the bold man
the bold and the beautiful.

Instead of a single element being responsible for the nominal group, it seems that a set of related factors is involved. For example, the determiner function may be unrealized if the noun is a plural or mass noun. The noun may be absent in the case of a combination of definite determiner and adjective, but this may only lead to plural interpretation: in the case of singular reference, or in the case of an indefinite article, the head noun must be present in the form of the pro-form one(s). In this way, various aspects of the elements making up a noun phrase can be found to be interrelated. Dependency-based modelling does not reflect this interrelatedness, and its analysis suggests that there is always a uniquely identifiable element responsible for certain properties. One consequence of this approach, as becomes clear from the above dilemma, is the necessity to determine a unique direction of the dependency relation between determiner and noun. In NLCA, this dilemma does not arise; moreover, there is a very natural way in which the different realizations of the nominal group follow from the interaction of the relations identified.6 There are various reasons for this.

First, NLCA identifies different kinds of relations and treats them differently. The qualification relation is essentially different from the predication relation, which results in the fact that only qualifiers, not predicates, may

---

6Cf. the discussion of Ex. 2 above, in particular, with respect to nominal adjective phrases and quantifying pronouns.
introduce Proto-items. This, as was explained above, is due to the fact that the qualifier does not introduce information that is independent of the core. The symmetrical instantiation of predication (major predication) is different from the asymmetrical instantiation (minor predication). This difference becomes manifest in the fact that in the latter, only the (minor) predicate may invoke the relation, whereas in the former, both participants invoke the relation, not just the (major) predicate. Note also that in the case of minor predicates, the functor is, in fact, optional (e.g. the adjective in an adjective-noun combination), which could never arise if functors were always taken to be the head of a construct. In this way NLCA distinguishes between the relative contribution that each type of relation makes with regard to the larger unit that is created. None of these factors is modelled in a purely dependency-based approach which operates on the basis of a single type of relation only. Moreover, in NLCA, items are typically involved in several relations simultaneously. This is not usually the case in dependency-based modelling (cf. the “No-tangling principle” [Hud95], see also [Fra96], which says that one dependency arrow points at every word. It is telling that this principle seems to be untenable for certain types of sentences.)

The following diagram reveals some differences between the directionality of the Head-Dependent relation of Word Grammar and the Functor-Argument relation of NLCA, generalising over the three different relations. (A → B is to be understood as: ‘A entails the presence of B’; in the dependency relation, this implies that the arrow points from the dependent element to the head. A ↔ B is equivalent to A → B and A ← B).

<table>
<thead>
<tr>
<th>Relation</th>
<th>WG</th>
<th>NLCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb - noun</td>
<td>H ← D</td>
<td>F ↔ A</td>
</tr>
<tr>
<td>noun - adjective</td>
<td>H ← D</td>
<td>A ← F</td>
</tr>
<tr>
<td>article - common noun</td>
<td>H ← D</td>
<td>F → A</td>
</tr>
</tbody>
</table>

A second characteristic of NLCA is that it distinguishes between internal and external arguments, which play a different role in the combinatorics of the model. It is the role of internal and external arguments and their relations that allows NLCA to model a situation where a particular property is distributed over a number of different elements (where ‘distributed over a number of elements’ is to be distinguished from ‘present at each of those elements’). In the case of the nominal group, for instance, this property concerns referential status. A noun phrase is used adequately in context if it permits identification of a referent, but whether or not it is able to do so
depends on a combination of factors; not on one in particular.

The role of internal and external arguments in NLCA derives from a property that dependency-based modelling does not use: the possibility that phrase structure units carry additional information beyond their constituent parts (cf. [Hud95] cited above). NLCA uses this information in the relations it identifies. By incorporating these relations as intrinsic parts of the units of analysis, it explains the formation of grammatical structures in a more fundamental way than dependency-based modelling.

A third, crucial factor that underlies the different results obtained in NLCA relates to the fact that the relations upon which the combinatorial mechanism of the model operates do not involve syntactic units per sé, but rather, build on conceptual distinctions underlying the concrete manifestation at the level of syntax, morphology etc. This is why NLCA is able to identify a wide range of possible instantiations of a noun phrase without special provisions. In order to fulfil its communicative role, a noun phrase, in our view, requires a certain referential clarity on the conceptual level. This referential clarity may be realized in different ways. There are different types of reference [QGLS85]: noun phrases may show variation in specificity (e.g. specific, generic or unique reference), definiteness (definite or indefinite), and number (singular, plural or mass). These types of referential information may be expressed at the syntactic level (e.g. by the presence of a determiner), the morphological level (e.g. a plural suffix) or in terms of a more abstract feature (e.g. MASS). Each type of reference has consequences for the potential realization of the determiner in English NPs. In NLCA, however, these will all be regarded as qualifiers entering into a Q-relation with their core. It is the Q-relation that gives the noun phrase its referential completeness, not a determiner or a head noun as such. This accounts for the following NP’s without the need to postulate any further structural requirements:

- the man \( Q \)-relation \( \text{the-man} \).
- men \( Q \)-relation \( \text{PLURAL-man} \).
- the bold man \( Q \)-relation \( \text{the-man} \);
  \( mp \)-relation \( \text{bold-man} \).
- the bold \( Q \)-relation \( \text{the-Proto-object} \);
  \( mp \)-relation \( \text{bold-Proto-object} \).
- sugar \( Q \)-relation \( \text{MASS-sugar} \).
- John \( Q \)-relation \( \text{UNIQUE-John} \).

There are also deviant instantiations of NPs, for example, with a determiner accompanying a proper noun (e.g. *That John is a teacher*) or an
empty determiner with a count singular noun (e.g. in PPs like *at school*). Interestingly, again no special provisions seem to be necessary in order to account for such cases. The demonstrative pronoun *that* may be analysed as an attribute that creates a Proto-object of which it is the internal argument; the Proto-object itself points to the external argument of the attribute (cf. the discussion of *some* in Ex. 2). *John*, which normally has unique reference, may replace the Proto-object, but then its qualifying internal argument must be unified with the demonstrative *that*, suggesting (correctly) that a specific John is involved. With bare nouns, as in *at school*, the qualifying internal argument of the noun is not realized, which means that it is not specified for its referential properties. However, this is precisely the nature of the difference between *at the school* and *at school*, where the latter relates to the institution (i.e. the phenomenon of school) rather than a specific external referent. This, then, follows from the treatment of qualifiers in general.

7 Summary and conclusion

We have presented a model that exemplifies how structure in natural language can be derived dynamically, based on the (combinatorial) properties of lexical items and a set of general principles. The advantage of such an approach lies in its flexibility that makes it applicable to difficult syntactic phenomena like coordination, discontinuous structures and structural variation. The combinatorial mechanism of NLCA operates on the basis of three relation schemes, based on conceptual rather than grammatical distinctions. Herein lie some of the differences with other word-based models such as dependency. An implementation of NLCA is currently under development, and applications to languages besides English (e.g. Hungarian) envisaged.

References


7The development of NLCA has up to this point been focused on analysis; application in production/generation may reveal that slightly more specific restrictions on co-occurrence apply. For example, if a definite determiner were used with a proper noun one would expect a further modification, e.g. in the form of a relative clause: *The John I saw was very tall*. The same may apply to other combinations of lexical units.


