An Algebraic Approach to the German Noun Phrase

Joachim Lambek, Anne Preller

An algebraic approach to the German noun phrase

Joachim Lambek
Department of Mathematics
MacGill University
Montréal, Québec, Canada
lambek@math.mcgill.ca

Anne Preller
LIRMM/ CNRS
Montpellier, France
France
preller@lirmm.fr

Abstract: We analyze the German noun phrase by assigning appropriate types to each noun, adjective and determiner, so that the type of the whole noun phrase can be checked by a calculation. In particular, the types will serve to distinguish between strong and weak declension of the adjective. The types are elements of the free pregroup generated by a partially ordered set; but very little of this mathematical apparatus is required for applications limited to the noun phrase.

How does a child growing up in Germany know to say "*der kleine Mann*", but "*ein kleiner Mann*"? Native German speakers will make the correct distinction, without being told that there is a difference between the strong and weak declension of the adjective. Evidently, speakers of German, or of any other language for that matter, carry out rapid subconscious calculations, employing internalized rules of grammar that they may not consciously be aware of. As far as we know, the weak versus strong distinction is peculiar to Germanic languages.

The purpose of this article is to demonstrate that agreement can be handled by the algebraic notion of pregroup [**1**]. It is part of an ongoing investigation of German sentence structure with the help of pregroups. This mathematical technique had been shown to be applicable to German word order [**2**] and is being further elaborated in an article in preparation [**3**]. Here we shall concentrate on a pregroup analysis of the German noun phrase. We are not trying to discover new rules of German grammar, on which a vast literature exists. Instead, we take the naïve approach of anthropologists: Looking at the input of a couple of native speakers and attempting to frame some simple rules that will explain this. Only when our native informants were unsure did we look at the standard reference work, the Duden [**4**].

As it turns out, for the purpose of studying the noun phrase only, most of the mathematical machinery is not needed. All the reader has to know is that we assign a *type,* which is a string of *simple types*, to each German word, here noun, adjective or determiner. Each simple type is either a basic type a or has the form a^{ℓ} or a^{r}. (Simple types of the form $a^{\ell\ell}$ and perhaps a^{rr} are needed elsewhere [e.g. **2**, **3**], but not here, and even simple types of the form a^{r} are only necessary for morphemes.)

The *basic types* are elements of a given partially ordered set. The order relation is denoted by an arrow and is assumed to be reflexive, transitive and anti-symmetric :

$$a \to a\,, \qquad \frac{a \to b \quad b \to c}{a \to c}\,, \qquad \frac{a \to b \quad b \to a}{a = b}\,.$$

The order relation is extended to simple types thus :

$$\frac{a \to b}{b^{\ell} \to a^{\ell}}\,, \qquad \frac{a \to b}{b^{r} \to a^{r}}$$

and to strings of simple types thus :

$$\frac{\alpha \to \beta \quad \gamma \to \delta}{\alpha\gamma \to \beta\delta}\,.$$

The only mathematical rules needed here are *contractions* :

$$a^{\ell}a \to 1\,, \qquad aa^{r} \to 1\,,$$

where 1 is the empty string.

A *noun phrase*, if it is not a name or a pronoun, may be complete or incomplete. An *incomplete* noun phrase must end in a noun and may be preceded by a string of adjectives. A *complete* noun phrase is formed from an incomplete one, by placing a determiner in front. In the present context, we may also think of pronouns as complete noun phrases. However, in different contexts, they may have to be distinguished from other noun phrases. Not only do they carry an additional feature, that of *person* (first, second, third), but they may occupy different positions in a sentence :

*ich gebe es ihm, *ich gebe das Buch ihm.*

Among the nouns, one must distinguish between *count nouns* and *mass nouns*. The former are inflected according to *gender* (masculine, feminine, neuter), *number* (singular, plural) and *case* (nominative, genitive, dative, accusative). We shall assume here that the inflected forms are stored in the mental dictionary and refrain from discussing whatever idiosyncratic rules may generate them. To each count noun, we assign a type $\boldsymbol{c}_{gnc}$, where $g = 1,2,3$ stands for the gender, $n = 1,2$ for the number and $c = 1,2,3,4$ for the case. The mass nouns also carry gender and case, but they have no plural; we assign to them the type $\boldsymbol{m}_{g1c}$. [1]

[1] Many nouns can switch categories. For example *Wein* is usually a mass noun, but it can be used as a count noun, when denoting a "kind of wine". In that role it also has a plural, as in

die französischen Weine.

On the other hand, it is difficult to think of a plural for *Obst.* Would it be *Obste* or *Öbster*? The information whether a noun is a count noun, a mass noun or both, is to be found in the dictionary.

Names require special consideration. Usually, they are already complete noun phrases, but they can also serve as count nouns, necessarily so, when preceded by an adjective or in the plural :

die faule Marie, *die Kennedys.*

The type of a complete noun phrase also encodes gender, number and case. At first sight, the gender of a complete noun phrase would seem irrelevant, but it may play a role in certain contexts, e.g.

(eine) der Frauen, *(keiner) der Männer*.

We assign the type $\boldsymbol{n}_{gnc}$ to a complete noun phrase.

An incomplete noun phrase also displays a fourth feature, its *strength*, which depends on the endings of any adjectives preceding it, the bane of all non-native learners of German. Thus, *gute Mann* and *guter Mann* are both correct incomplete noun phrases, to be completed as : *der gute Mann* and *ein guter Mann.* Moreover, the type of an incomplete noun phrase must contain the information whether the noun at its extreme right is a count noun or a mass noun. We assign to an incomplete noun phrase the type $\boldsymbol{c}_{sgnc}$ or $\boldsymbol{m}_{sg1c}$, where *s =1* or *2* depending whether it is *strong* or *weak*.

Since a naked noun bears no sign of strength, we postulate

$$\boldsymbol{x}_{gnc} \rightarrow \boldsymbol{x}_{sgnc} \qquad (\boldsymbol{x} = \boldsymbol{m} \text{ or } \boldsymbol{c}\ ).$$

(Recall that the basic types are elements of a partially ordered set.) Since mass nouns and plural count nouns do not necessarily require a determiner, we also postulate

$$\boldsymbol{m}_{sg1c} \rightarrow \boldsymbol{n}_{g1c} \qquad\qquad \boldsymbol{c}_{sg2c} \rightarrow \boldsymbol{n}_{g2c}\,.$$

However, we must emphasize that

$$\boldsymbol{c}_{sg1c} \not\rightarrow \boldsymbol{n}_{g1c}\,.$$

The canonical way to convert an incomplete noun phrase to a complete one is with the help of a definite article. This article may be displayed by a four-by-four matrix :

$$D = \begin{pmatrix} der & des & dem & den \\ die & der & der & die \\ das & des & dem & das \\ die & der & den & die \end{pmatrix}$$

Here, the first three rows correspond to masculine, feminine, neuter singular and the fourth row contains the four cases of the plural, independent of gender. We let $g' = 1,2,3,4$, so the entries of this matrix may be denoted by $D_{g'c}$. We assign the following types to the definite article :

$$D_{4c} \ : \ \boldsymbol{n}_{g2c}\, \boldsymbol{c}_{2g2c}{}^{\ell} \qquad (\, g = 1,2,3\,,\ c = 1,2,3,4\,),$$

$$D_{gc} \ : \ \boldsymbol{n}_{g1c}\boldsymbol{x}_{2g1c}{}^{\ell} \qquad (\, \boldsymbol{x} = \boldsymbol{m} \text{ or } \boldsymbol{c}\,,\ g = 1,2,3\,).$$

This will ensure that an incomplete noun phrase following a definite article must be weak.

The demonstrative determiners are best seen as a stem *dies* or *jen* followed by a *very strong* ending $E_{g'c}$, given by the matrix

$$E \ = \begin{pmatrix} er & es & em & en \\ e & er & er & e \\ es & es & em & es \\ e & er & en & e \end{pmatrix}$$

The endings are essentially those of the definite article, except that we must replace *ie* by *e* and *as* by *es*. The types of $diesE_{g'c}$ and $jenE_{g'c}$ are the same as those of $D_{g'c}$.

The quantifier *kein* and the possessive determiners *mein, dein, uns(e)r* etc. also take very strong endings, except that in three places, $(g',c) = (1,1),(3,1),(3,4)$, the ending, call it $E'_{g'c}$, is empty. Thus

$$E' \ = \begin{pmatrix} \varnothing & es & em & en \\ e & er & er & e \\ \varnothing & es & em & \varnothing \\ e & er & en & e \end{pmatrix}.$$

When $E'_{g'c} = \varnothing$, any adjective following it must be strong. Otherwise $keinE'_{g'c}$ etc. has the same type as $D_{g'c}$ and requires the subsequent adjective to be weak. Thus, when $\boldsymbol{x} = \boldsymbol{m}$ or $\boldsymbol{c}$,

$$keinE'_{g'c} \ : \ \begin{cases} \boldsymbol{n}_{g'1c}\, \boldsymbol{x}_{1g'1c}{}^{\ell} & \text{if } (g',c) = (1,1),\,(3,1) \text{ or } (3,4) \\ \boldsymbol{n}_{g'nc}\, \boldsymbol{x}_{2g'nc}{}^{\ell} & \text{otherwise} \end{cases}.$$

The indefinite article *ein* takes the same endings as *kein*, except that the bottom row of the matrix, representing the plural, is missing and its type requires $\boldsymbol{x} = \boldsymbol{c}$, as it cannot modify a mass noun. [2] Thus, its inflected forms are typed as follows :

$$einE'_{gc} \ : \begin{cases} \boldsymbol{n}_{g1c}\boldsymbol{c}_{1g1c}{}^{\ell} & \text{if } \boldsymbol{(}g,c\boldsymbol{)} = \boldsymbol{(}1,1\boldsymbol{),(}3,1\boldsymbol{),(}3,4\boldsymbol{)} \\ \boldsymbol{n}_{g1c}\boldsymbol{c}_{2g1c}{}^{\ell} & \text{otherwise} \end{cases}.$$

Now let us turn to the adjective, say A. When used predicatively, it is uninflected; but, when used attributively in a noun phrase, it takes the form $AE_{sg'c}$ with two declensions, a strong one ($s = 1$) and a weak one ($s = 2$). The endings $E_{sg'c}$ are given by two matrices :

$$E_1 = \begin{pmatrix} er & en & em & en \\ e & er & er & e \\ es & en & em & es \\ e & er & en & e \end{pmatrix},$$

$$E_2 = \begin{pmatrix} e & en & en & en \\ e & en & en & e \\ e & en & en & e \\ en & en & en & en \end{pmatrix}.$$

Note that the strong endings are the same as the very strong ones, except that in two places in the second column the weak ending is substituted for the strong one.

We assign types to the inflected forms of the adjectives as follows :

$$AE_{sgc} \ \boldsymbol{:} \ \boldsymbol{x}_{sg1c}\boldsymbol{x}_{sg1c}{}^{\ell}, \ g = 1,2,3, \ \boldsymbol{x} = \boldsymbol{c} \text{ or } \boldsymbol{m}$$

$$AE_{s4c} \ \boldsymbol{:} \ \boldsymbol{x}_{sg2c}\boldsymbol{x}_{sg2c}{}^{\ell}, \ g = 1,2,3, \ \boldsymbol{x} = \boldsymbol{c} \ .$$

The adjective converts one incomplete noun phrase into another one of the same strength. This implies that consecutive adjectives have either both a weak ending or both a strong ending, they have *parallel* declension.

Assuming that a predicative adjective A has type $\boldsymbol{a}$, we could also assign types to the entries of the matrix E_s as follows :

---

[2] In the noun phrase *ein guter Wein*, the meaning of *Wein* is that of a kind of wine and its type is that of a count-noun.

E_{sgc} **:** $\boldsymbol{a}^r \boldsymbol{x}_{sg1c} \boldsymbol{x}_{sg1c}{}^{\ell}$

E_{s4c} **:** $\boldsymbol{a}^r \boldsymbol{x}_{sg2c} \boldsymbol{x}_{sg2c}{}^{\ell}$.

This is the only time where the right adjoint of a basic type would be required.

In the following examples, we recall that $\boldsymbol{x}_{gnc} \to \boldsymbol{x}_{sgnc}$, $\boldsymbol{x} = \boldsymbol{c}$ or $\boldsymbol{m}$, so that

$$\boldsymbol{x}_{sgnc}{}^{\ell} \boldsymbol{x}_{gnc} \to 1 .$$

EXAMPLES :

die schöne Frau

$$(\boldsymbol{n}_{21c} \underbrace{\boldsymbol{c}_{221c}{}^{\ell})(\boldsymbol{c}_{221c}}\ \underbrace{\boldsymbol{c}_{221c}{}^{\ell})\boldsymbol{c}_{21c}} \to \boldsymbol{n}_{21c}, \qquad (c = 1,4)$$

der schönen Frauen

$$(\boldsymbol{n}_{222} \underbrace{\boldsymbol{c}_{2222}{}^{\ell})(\boldsymbol{c}_{2222}}\ \underbrace{\boldsymbol{c}_{2222}{}^{\ell})\boldsymbol{c}_{222}} \to \boldsymbol{n}_{222},$$

keiner schönen guten Frau

$$(\boldsymbol{n}_{21c} \underbrace{\boldsymbol{c}_{221c}{}^{\ell})(\boldsymbol{c}_{221c}}\ \underbrace{\boldsymbol{c}_{221c}{}^{\ell})(\boldsymbol{c}_{221c}}\ \underbrace{\boldsymbol{c}_{221c}{}^{\ell})\boldsymbol{c}_{21c}} \to \boldsymbol{n}_{21c}, \qquad (c = 2,3)$$

(mit) gutem Wein

$$(\boldsymbol{m}_{1113} \underbrace{\boldsymbol{m}_{1113}{}^{\ell})}\ \boldsymbol{m}_{113} \to \boldsymbol{m}_{1113} \to \boldsymbol{n}_{113}$$

(mit) diesem guten Wein

$$(\boldsymbol{n}_{113} \underbrace{\boldsymbol{m}_{2113}{}^{\ell})}\ (\boldsymbol{m}_{2113} \underbrace{\boldsymbol{m}_{2113}{}^{\ell})}\ \boldsymbol{m}_{113} \to \boldsymbol{n}_{113}$$

ein frohes glückliches Kind

$$(\boldsymbol{n}_{31c} \underbrace{\boldsymbol{c}_{131c}{}^{\ell})}\ (\boldsymbol{c}_{131c} \underbrace{\boldsymbol{c}_{131c}{}^{\ell})}\ (\boldsymbol{c}_{131c} \underbrace{\boldsymbol{c}_{131c}{}^{\ell})}\ \boldsymbol{c}_{31c} \to \boldsymbol{n}_{31c}, \qquad (c = 1,4)$$

The so-called “indefinite” count adjectives have an ambiguous status, between determiner and adjective. The Duden [**4**] distinguishes sixteen different combinations. Each of them can be reflected by appropriate typing which we shall demonstrate for the quantifier *viel*. As a determiner, it has no ending and modifies mass nouns:

$viel \;:\; \boldsymbol{n}_{g1c}\boldsymbol{m}_{1g1c}{}^{\ell}$

EXAMPLES :

viel *guter* *Wein*

$(\boldsymbol{n}_{111}\ \underbrace{\boldsymbol{m}_{1111}{}^{\ell})\ (\boldsymbol{m}_{1111}}\ \underbrace{\boldsymbol{m}_{1111}{}^{\ell})\ \boldsymbol{m}_{111}} \rightarrow \boldsymbol{n}_{111}$

(mit) *viel* *unendlicher Geduld*

$(\boldsymbol{n}_{213}\ \underbrace{\boldsymbol{m}_{1213}{}^{\ell})\ (\boldsymbol{m}_{1213}}\ \underbrace{\boldsymbol{m}_{1213}{}^{\ell})\ \boldsymbol{m}_{213}} \rightarrow \boldsymbol{n}_{213}.$

But *viel* can also occur as an adjective, although it then must precede all other adjectives (see below). As an adjective, *viel* modifies mass nouns in the singular and count nouns in the plural. It has the same endings as the other adjectives. With the exception of the strong ending nominative, dative and accusative neuter and dative masculine, the type of $vielE_{sg'c}$ is the same as that of ordinary adjectives which have parallel declension:

der viele gute Wein; vieler guter Wein,
die vielen schönen Blumen; viele schöne Blumen
das viele dumme Geschwätz
(mit) dem vielen dummen Geschwätz
(mit) dem vielen guten Wein

In the nominative, dative and accusative neuter and in the dative masculine, however, *viel* with a strong ending behaves like a demonstrative determiner. Hence the following adjective must have a weak ending:

vieles dumme Geschwätz
(mit) vielem dummen Geschwätz
(mit) vielem guten Wein

Therfore, in the nominative, dative and accusative neuter and in the dative masculine, *viel* with a strong ending has the following types

$vielE_{13c} \;:\; \boldsymbol{m}_{131c}\boldsymbol{m}_{231c}{}^{\ell},\ c = 1,3,4$

$vielE_{113} \;:\; \boldsymbol{m}_{1113}\boldsymbol{m}_{2113}{}^{\ell}.$

Another problem is that, as an adjective, *viel* must precede all other adjectives. In fact, the Duden [**1**] considers four groups of adjectives, as exemplified in :

die vielen damaligen unangenehmen steuerlichen Probleme.

1 2 3 4

Adjectives of group 1 must precede those of group 2, etc., but adjectives in the same group are interchangeable, as in :

kalte, nasse Luft / nasse, kalte Luft.

As a rule, only adjectives in group 3 can be used predicatively, e.g.

**die Probleme sind steuerlich / damalig / viel*

though there are exceptions. To handle the problem posed by the four groups of adjectives, we would have to replace the single type $\boldsymbol{a}$ by four types $\boldsymbol{a}_1,...,\boldsymbol{a}_4$. This would necessitate some obvious modifications of our grammar, which we shall skip.

Adjectives of group 1 are also called "indefinite count adjectives". They tend to oscillate between being adjectives and being determiners. The rules vary from one count adjective to another (see [**4**], 471 – 473). To keep this article within reasonable bounds, we have confined discussion here to the indefinite count adjective *viel* (see above). For the same reason, we have omitted the discussion of numerals and nouns in the genitive acting as determiners.

Conclusion.

We have shown that the German noun phrase is constructed according to strict rules that can be formalized mathematically. The mathematical machinery we have chosen is that of a pregroup grammar, which has been used successfully in the analysis of a number of European languages. For the purpose of studying the noun phrase, a rather rudimentary form of pregroup grammar suffices. It could be easily replaced by a more traditional Adjukievicz-Bar Hillel grammar, provided only that the usual two-element set $\{\boldsymbol{S},\boldsymbol{N}\}$ of basic types (as in [**5**]) is extended to a much larger partially ordered set. Evidently, even less mathematized forms of grammar can be used for analyzing the noun phrase, as was done successfully in [**6**].

While the mathematical machinery we have employed is rather simple, compared with what is required for analyzing the sentence structure [**1**, **2**], e.g. double adjoints are not needed, there is an additional complication due to the many subscripts placed on the basic types. This complication is not needed for English, but arises whenever one looks at highly inflected languages like Latin or Russian.

REFERENCES

1. J. Lambek, Type grammars revisited, in : A. Lecomte et al., Logical Aspects of Computational Linguistics, Springer LNAI (1999), 1-27.

2. J. Lambek, Type grammar meets German word order, Theoretical Linguistics 26 (2000), 19-30.

3. J. Lambek, A. Preller, An algebraic approach to the German sentence, to appear

4. A. Klosa et al. (eds.), Duden 4, Die Grammatik, Dudenverlag, Mannheim etc. 1998

5. T.T. Ballmer, Sprachrekonstruktionssysteme, Kronberg; Scriptor (1975).

6. A. Zwicky, German adjective agreement in GSPG* , Linguistics 24 (1986), 957-990.