An algebraic approach to French sentence structure

D. Bargelli and J. Lambek*

McGill University, Montreal

We propose to investigate the structure of French sentences with the help of a minimal algebraic technique based on the assignment of types to words. Mathematically speaking, the types are elements of the free "pregroup" (a generalization of a partially ordered group) generated by a partially ordered set of basic types. In particular, this partial order is carefully adjusted to account for the order of preverbal clitic pronouns in a sentence.

1. Types.

The main idea is this: to each French word there are assigned one or more types so that the sentencehood of a string of French words can be checked by a simple calculation. To begin with, there are a number of *basic types* such as the following:

$\mathbf{s}_1$ for direct statements, that is, declarative sentences, in the present tense;
$\mathbf{s}_2$ for direct statements in the past (imperfect) tense;
$\mathbf{s}$ for direct statements when the tense does not matter;
$\overline{\mathbf{s}}$ for indirect statements.

The set of basic types is *partially ordered* by a relation $\to$. By this is meant a binary relation satisfying the following axioms and rules of inference:

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

Furthermore we shall postulate that e.g.

$$\mathbf{s}_1 \to \mathbf{s}, \quad \mathbf{s}_2 \to \mathbf{s}, \quad \mathbf{s} \to \overline{\mathbf{s}}.$$

In fact, we shall adopt the convention that $a \to \overline{a} \to \overline{\overline{a}}$ for any basic type a. The bar here plays a rôle similar to that in the $\overline{X}$-theory of Chomsky and Jackendoff [1977]; but for us it is merely a notational device, not driven by their theory.

From the basic types we construct *simple types*: if a is a simple type, then so are a^ℓ and a^r, called the *left* and *right adjoint* respectively. Thus, if a is a basic type then

$$a,\ a^\ell,\ a^{\ell\ell}, \cdots, a^r,\ a^{rr}, \cdots$$

are simple types.

By a *type* we shall mean a string of simple types $a_1 a_2 \cdots a_n$. In particular, if $n = 1$, this implies that a simple type is a type; and, if $n = 0$, according to the usual mathematical convention, the empty string 1 is a type, it being understood that

$$a1 = a = 1a.$$

The partial order $\to$ may be extended by the rule

$$\frac{a \to a' \quad b \to b'}{ab \to a'b'}$$

to the monoid of all types. Moreover, we postulate the following *contraction rules*:

$$a^\ell a \to 1, \quad aa^r \to 1.$$

For the purpose of sentence verification these suffice; but mathematicians will also require the *expansion rules*:

$$1 \to aa^\ell, \quad 1 \to a^r a,$$

which will assure the uniqueness of the adjoints. For example, one can then prove that

$$a^{r\ell} = a, \quad a^{\ell r} = a,$$

but neither $a^{\ell\ell} = a$ nor $a^{rr} = a$. The adjoints may easily be extended to all types, that is, strings of simple types, by defining

$$1^\ell = 1,\ 1^r = 1,\ (ab)^\ell = b^\ell a^\ell,\ (ab)^r = b^r a^r.$$

2. Infinitives.

Crucial to all sentences are the verbs, usually represented by their infinitives. For example, we have

dormir of type $\mathbf{i}$
prendre of type $\mathbf{io}^\ell$
manger of type $\mathbf{i}$ or $\mathbf{io}^\ell$.

Here $\mathbf{i}$ and $\mathbf{o}$ are basic types:

$\mathbf{i}$ for infinitives of intransitive verbs,
$\mathbf{o}$ for direct objects (COD).

For example,

$$\begin{array}{cc} \textit{manger} & \underbrace{\textit{une pomme}} \\ (\mathbf{io}^\ell) & \mathbf{o} \end{array}$$

is an expression of type

$$(\mathbf{io}^\ell)\mathbf{o} = \mathbf{i}(\mathbf{o}^\ell\mathbf{o}) \to \mathbf{i}1 = \mathbf{i}.$$

Direct objects are only one kind of verb complements. Others are indirect objects (CID) and locatives, as in

$$\begin{array}{cc} \textit{obéir} & \underbrace{\textit{à Jean}} \\ (\mathbf{i}\omega^\ell) & \omega \end{array}$$

and

$$\begin{array}{cc} \textit{habiter} & \underbrace{\textit{à Paris}} \\ (\mathbf{i}\lambda^\ell) & \lambda \end{array} .$$

We have used types ω and λ as follows:

ω for indirect objects,
λ for locatives.

3. Noun-phrases.

If separate types for *une, pomme, Jean, Paris* and *à* are required, we introduce new basic types:

c for count nouns, e.g. *pomme*,
$\mathbf{n}_s$ for singular noun-phrases, e.g. *Jean, Paris*.

Hence we are led to assign

$\mathbf{n}_s\mathbf{c}^\ell$ to the article *une*,
$\omega\mathbf{o}^\ell$ and $\lambda\ \mathbf{o}^\ell$ to the preposition *à*.

We postulate $\mathbf{n}_s \to \mathbf{o}$ to indicate that singular noun-phrases may be direct objects. Thus, we may calculate the types of *une pomme, à Jean* and *à Paris* as follows:

$$\begin{array}{rl} (\mathbf{n}_s\mathbf{c}^\ell)\mathbf{c} = & \mathbf{n}_s(\mathbf{c}^\ell\mathbf{c}) \to \mathbf{n}_s 1 = \mathbf{n}_s, \\ (\omega\mathbf{o}^\ell)\mathbf{n}_s \to & \omega(\mathbf{o}^\ell\mathbf{o}) \to \omega 1 = \omega, \\ (\lambda\mathbf{o}^\ell)\mathbf{n}_s \to & \lambda(\mathbf{o}^\ell\mathbf{o}) \to \lambda 1 = \lambda. \end{array}$$

For the last two calculations we recall that $n_s \to \mathbf{o}$.

For later use we also mention the following basic types:

m for mass nouns, e.g. *pain*,
p for plurals (usually of count nouns), e.g. *pommes*,
$\mathbf{n}_p$ for plural noun-phrases.

We postulate $\mathbf{n}_p \to \mathbf{o}$ to indicate that plural noun-phrases can also occur as direct objects. We can now account for such noun-phrases as *du pain* and *des pommes* with the help of the indefinite articles

du of type $\mathbf{n}_s\mathbf{m}^\ell$,
des of type $\mathbf{n}_p\mathbf{p}^\ell$.

We have ignored here one essential fact of French grammar, namely that *pomme* and *une* are both feminine and that their genders must agree. But, in this first attempt to apply our algebraic technique to French grammar, we choose to ignore some complications, just as Galileo, in his first attempt to analyze motion mathematically, chose to ignore friction.

4. Extended infinitives.

In English, the noun phrase *an apple* can be replaced by the pronoun *it* in the same position. In French, the situation is more complicated: the clitic pronoun *la* appears before

the verb, as in *la + manger*. We want this expression to be treated like an infinitive of an intransitive verb, so we should assign to it the type $\mathbf{i}$. For reasons that will become clear later, we assign to it the type $\bar{\mathbf{i}}$ instead, subject of course to the rule $\mathbf{i} \to \bar{\mathbf{i}}$. We accomplish this by assigning to the clitic pronoun *la* the type $\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell}$, where, for reasons that will be discussed later, we have put a bar on the second $\mathbf{i}$ as well, so *la + manger* has type

$$\begin{array}{ll} (\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell}) = \bar{\mathbf{i}}(\mathbf{o}^{\ell\ell}(\bar{\mathbf{i}}^{\ell}\mathbf{i})\mathbf{o}^{\ell}) & \\ \to\ \bar{\mathbf{i}}(\mathbf{o}^{\ell\ell}\mathbf{o}^{\ell}) & \text{since } \bar{\mathbf{i}}^{\ell}\mathbf{i} \to \bar{\mathbf{i}}^{\ell}\bar{\mathbf{i}} \to 1 \\ \to\ \bar{\mathbf{i}} & \text{since } \mathbf{o}^{\ell\ell}\mathbf{o}^{\ell} \to 1 \end{array}$$

We represent this calculation diagramatically, by a method that goes back to Z. Harris [1966], as follows:

$$\begin{array}{c} la\ +\ manger \\ \underbrace{(\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell})\quad (\mathbf{i}\mathbf{o}^{\ell})} \end{array}$$

Note that we have inserted the symbol + to prepare the reader for our claim that *la + manger* is to be treated like a single word, we shall call it an extended infinitive of type $\bar{\mathbf{i}}$.

We recall that there are other kinds of direct objects, constructed from mass nouns or plurals, such as *du pain* of type $\mathbf{n}_s \to \mathbf{o}$ and *des pommes* of type $\mathbf{n}_p \to \mathbf{o}$, with the help of the indefinite articles *du* and *des*. They are represented by another preverbal clitic pronoun *en*, as in *en + manger*, another extended infinitive, this time of type $\mathbf{i}$. We accomplish this by assigning to *en* the type $\mathbf{i}\mathbf{o}^{\ell\ell}\mathbf{i}^{\ell}$ without bars.

Consider now a verb, such as *donner*, which requires two objects, a direct one and an indirect one, as in

donner une pomme à Jean,

donner à Jean une pomme.

The second of these two sentences is less common, but it is permitted for emphasis. Recalling that the indirect object has type ω, we require that *donner* has two types, namely $\mathbf{i}\omega^{\ell}\mathbf{o}^{\ell}$ and $\mathbf{i}\mathbf{o}^{\ell}\omega^{\ell}$. For example, we have

$$\begin{array}{c} donner\ \underbrace{une\ pomme}\ \underbrace{à\ Jean} \\ \underbrace{(\mathbf{i}\omega^{\ell}\mathbf{o}^{\ell})\quad \mathbf{o}\quad \omega} \end{array}$$

The indirect object *à Jean* may be replaced by the preverbal pronoun *lui* of type $\bar{\mathbf{i}}\omega^{\ell\ell}\mathbf{i}^{\ell}$, to justify the following:

$$\begin{array}{c} lui\ +\ donner\ \underbrace{la\ pomme}, \\ \underbrace{(\bar{\mathbf{i}}\omega^{\ell\ell}\mathbf{i}^{\ell})}\underbrace{(\mathbf{i}\omega^{\ell}}\underbrace{\mathbf{o}^{\ell})}\mathbf{o} \end{array}$$

$$la\ +\ donner\ \underbrace{à\ Jean}, \qquad (\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell})(\mathbf{io}^{\ell}\omega^{\ell})\omega$$

$$en\ +\ donner\ \underbrace{à\ Jean}, \qquad (\mathbf{io}^{\ell\ell}\mathbf{i}^{\ell})(\mathbf{io}^{\ell}\omega^{\ell})\omega$$

$$la\ +\ lui\ +\ donner, \qquad (\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell})(\bar{\mathbf{i}}\omega^{\ell\ell}\mathbf{i}^{\ell})(\mathbf{i}\omega^{\ell}\mathbf{o}^{\ell})$$

$$lui\ +\ en\ +\ donner, \qquad (\bar{\mathbf{i}}\omega^{\ell\ell}\mathbf{i}^{\ell})(\mathbf{io}^{\ell\ell}\mathbf{i}^{\ell})(\mathbf{io}^{\ell}\omega^{\ell})$$

However, we are forbidden to say

$$*\ lui\ +\ la\ +\ donner, \qquad (\bar{\mathbf{i}}\omega^{\ell\ell}\mathbf{i}^{\ell})(\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell})(\mathbf{io}^{\ell}\omega^{\ell})$$

$$*\ en\ +\ lui\ +\ donner, \qquad (\mathbf{io}^{\ell\ell}\mathbf{i}^{\ell})(\bar{\mathbf{i}}\omega^{\ell\ell}\mathbf{i}^{\ell})(\mathbf{i}\omega^{\ell}\mathbf{o}^{\ell})$$

In fact, it was to avoid these contractions that the bars were introduced. Note that $a^{\ell}\overline{a} \not\to 1$, but

$$\overline{a}^{\ell}a \to \overline{a}^{\ell}\overline{a} \to 1.$$

5. Other clitic pronouns.

The clitic pronouns *la, lui* and *en* are not the only ones:

the accusative pronouns *le, la, les* have type $\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell}$;
the dative pronouns *lui, leur* have type $\bar{\mathbf{i}}\omega^{\ell\ell}\mathbf{i}^{\ell}$;
the partitive pronoun *en* has type $\mathbf{io}^{\ell\ell}\mathbf{i}^{\ell}$

the personal pronouns *me, te, se, nous* and *vous* can be either accusative of type $\overline{=}\mathbf{i}\,\mathbf{o}^{\ell\ell}\mathbf{i}^{\ell}$ or dative of type $\overline{=}\mathbf{i}\,\omega^{\ell\ell}\bar{\mathbf{i}}^{\ell}$. (The last two are not strictly speaking clitics, but they should be treated the same way.) Why the bars?

We want to admit

$$vous\ +\ les\ +\ offrir, \qquad (\overline{=}\mathbf{i}\,\omega^{\ell\ell}\bar{\mathbf{i}}^{\ell})(\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell})(\mathbf{io}^{\ell}\omega^{\ell})$$

but not

$$*\ les\ +\ vous\ +\ offrir, \qquad (\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\mathbf{i}^{\ell})(\overline{=}^{\ell}\mathbf{i}\,\omega^{\ell\ell}\bar{\mathbf{i}}^{\ell})(\mathbf{i}\omega^{\ell}\mathbf{o}^{\ell})$$

We also want to avoid the following combinations:

$$\begin{array}{cc} *\ vous\ +\ lui, & *\ lui\ +\ vous, \\ (\bar{=}\mathbf{i}\mathbf{o}^{\ell\ell}\underline{\mathbf{i}^{\ell}})(\bar{\mathbf{i}}\omega^{\ell\ell}\mathbf{i}^{\ell}) & (\bar{\mathbf{i}}\omega^{\ell\ell}\underline{\mathbf{i}^{\ell}})(\bar{=}\mathbf{i}\mathbf{o}^{\ell\ell}\mathbf{i}^{\ell}) \end{array}$$

$$\begin{array}{cc} *\ vous\ +\ nous, & *\ vous\ +\ nous, \\ (\bar{=}\mathbf{i}\mathbf{o}^{\ell\ell}\underline{\bar{\mathbf{i}}^{\ell}})(\bar{=}\mathbf{i}\omega^{\ell\ell}\mathbf{i}^{\ell}) & (\bar{=}\mathbf{i}\omega^{\ell\ell}\underline{\bar{\mathbf{i}}^{\ell}})(\bar{=}\mathbf{i}\mathbf{o}^{\ell\ell}\mathbf{i}^{\ell}) \end{array}$$

Finally, there is also the locative clitic y of type $\bar{\mathbf{i}}\lambda^{\ell\ell}\mathbf{i}^{\ell}$. Consider

$$\begin{array}{cc} aller & \underbrace{\grave{a}\quad Paris} \\ (\mathbf{i}\underline{\lambda^{\ell}}) & \lambda \end{array}$$

where λ is the type of a locative expression, and

$$\begin{array}{ccc} y & + & aller \\ (\bar{\mathbf{i}}\lambda^{\ell\ell}\mathbf{i}^{\ell}) & & (\mathbf{i}\lambda^{\ell}) \end{array}$$

The first bar on the type of y will be justified later. Consider next *mettre* of type $\mathbf{i}\lambda^{\ell}\mathbf{o}^{\ell}$ or $\mathbf{i}\mathbf{o}^{\ell}\lambda^{\ell}$ as exemplified by

$$mettre \quad \underbrace{une\ pomme} \quad \underbrace{sur\ la\ table} \quad (\mathbf{i}\lambda^{\ell}\mathbf{o}^{\ell}) \qquad \mathbf{o} \qquad \lambda$$

or, for emphasis, by

$$\begin{array}{ccc} mettre & \underbrace{sur\ la\ table} & \underbrace{une\ pomme} \\ & (\mathbf{i}\mathbf{o}^{\ell}\lambda^{\ell}) \qquad \lambda & \mathbf{o} \end{array}$$

The clitic y then appears in

$$\begin{array}{c} y\ +\ mettre\ \underbrace{une\ pomme}, \\ (\bar{\mathbf{i}}\lambda^{\ell\ell}\mathbf{i}^{\ell})(\mathbf{i}\lambda^{\ell}\mathbf{o}^{\ell})\mathbf{o} \end{array}$$

$$\begin{array}{c} la\ +\ y\ +\ mettre, \\ (\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\underline{\bar{\mathbf{i}}^{\ell}})(\bar{\mathbf{i}}\lambda^{\ell\ell}\mathbf{i}^{\ell})(\mathbf{i}\lambda^{\ell}\mathbf{o}^{\ell}) \end{array}$$

$$\begin{array}{c} y\ +\ en\ +\ mettre, \\ (\bar{\mathbf{i}}\lambda^{\ell\ell}\mathbf{i}^{\ell})(\mathbf{i}\mathbf{o}^{\ell\ell}\mathbf{i}^{\ell})(\mathbf{i}\mathbf{o}^{\ell}\lambda^{\ell}) \end{array}$$

but not in

$$\begin{array}{c} *\ en\ +\ y\ +\ mettre, \\ (\mathbf{i}\mathbf{o}^{\ell\ell}\underline{\mathbf{i}^{\ell}})(\bar{\mathbf{i}}\lambda^{\ell\ell}\mathbf{i}^{\ell})(\mathbf{i}\lambda^{\ell}\mathbf{o}^{\ell}) \end{array}$$

We shall list a few verbs together with some of their types and past participles. However, we will ignore the gender and the number of the latter, so that *dormi* represents *dormi(e)(s)*. A more complete list could be elaborated from Gross [1969], [Boons et al, 1972] and [Guillet and Leclere 1992]. We distinguish between past participles of type $\mathbf{p}_2$ and those of type $\mathbf{p}_2'$. The former require the auxiliary verb *avoir*, the latter *être*. We also use $\mathbf{a}$ for the type of adjectives.

dormir:	$\mathbf{i}$	*dormi*:	$\mathbf{p}_2$
venir:	$\mathbf{i}$	*venu*:	$\mathbf{p}_2'$
prendre:	$\mathbf{io}^\ell$	*pris*:	$\mathbf{p}_2\mathbf{o}^\ell$
manger:	$\mathbf{i}, \mathbf{io}^\ell$	*mangé*:	$\mathbf{p}_2, \mathbf{p}_2\mathbf{o}^\ell$
obéir:	$\mathbf{i}\omega^\ell$	*obéi*:	$\mathbf{p}_2\omega^\ell$
aller:	$\mathbf{i}\lambda^\ell$	*allé*:	$\mathbf{p}_2\lambda^\ell$
donner:	$\mathbf{i}\omega^\ell\mathbf{o}^\ell,\ \mathbf{io}^\ell\omega^\ell$	*donné*:	$\mathbf{p}_2\omega^\ell\mathbf{o}^\ell,\ \mathbf{p}_2\mathbf{o}^\ell\omega^\ell$
mettre:	$\mathbf{i}\lambda^\ell\mathbf{o}^\ell,\ \mathbf{io}^\ell\lambda^\ell$	*mis*:	$\mathbf{p}_2\lambda^\ell\mathbf{o}^\ell,\ \mathbf{p}_2\mathbf{o}^\ell\lambda^\ell$
vouloir:	$\mathbf{i}\overset{=\ell}{\mathbf{i}}$	*voulu*:	$\mathbf{p}_2\overset{=\ell}{\mathbf{i}}$
avoir:	$\mathbf{io}^\ell,\ \mathbf{ip}_2^\ell$	*eu*:	$\mathbf{p}_2\mathbf{o}^\ell,\ \mathbf{p}_2\mathbf{p}_2^\ell$
être:	$\mathbf{ia}^\ell,\ \mathbf{ip}_2'^\ell,\ \mathbf{io}^{\ell\ell}\mathbf{p}_2^\ell$	*été*:	$\mathbf{p}_2\mathbf{a}^\ell, {}^*\mathbf{p}_2\mathbf{p}_2'^\ell,\ \mathbf{p}_2\mathbf{o}^{\ell\ell}\mathbf{p}_2^\ell.$

6. Modal and auxiliary verbs.

The last three verbs in the above list require some discussion. The modal verb *vouloir*, like *pouvoir* and *devoir*, has been given type $\overset{=\!=^{\ell}}{\mathbf{i}\ \mathbf{i}}$ to avoid

$$*\ la\ +\ vouloir\ prendre,$$
$$(\underline{\mathbf{i}}\mathbf{o}^{\ell\ell}\underline{\bar{\mathbf{i}}}^{\ell})(\overset{=\!=^{\ell}}{\mathbf{i}\ \mathbf{i}})(\mathbf{io}^{\ell})$$

but to admit

$$vouloir\ la\ +\ prendre$$
$$(\overset{=\!=^{\ell}}{\mathbf{i}\ \mathbf{i}})\ (\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell})(\mathbf{io}^{\ell})$$

and

$$vouloir\ pouvoir\ venir$$
$$(\overset{=\!=^{\ell}}{\mathbf{i}\ \mathbf{i}})\ (\overset{=\!=^{\ell}}{\mathbf{i}\ \mathbf{i}})\quad \mathbf{i}$$

even

$$vouloir\ vouloir\ venir.$$

The verb *avoir* may occur as an ordinary transitive verb, as in

$$avoir\ \overbrace{une\ pomme}$$
$$\mathbf{io}^{\ell}\quad\ \mathbf{o}$$

but we are here interested in its rôle as an auxiliary verb to form the composite past, as in

$$avoir\quad dormi$$
$$(\mathbf{ip}_2^{\ell})\qquad\quad \mathbf{p}_2$$

or even in

$$avoir\ eu\ dormi,$$
$$(\mathbf{ip}_2^{\ell})\quad (\mathbf{p}_2\mathbf{p}_2^{\ell})\quad p_2$$

the so-called super-composite past, supposedly common in French Switzerland. Unfortunately, this type assignment also allows

$$avoir\ eu\ eu\ dormi,$$
$$(\mathbf{ip}_2^{\ell})\ (\mathbf{p}_2\mathbf{p}_2^{\ell})\ (\mathbf{p}_2\mathbf{p}_2^{\ell})\ \mathbf{p}_2$$

which should perhaps be ruled out on other grounds.

Care should be taken in analyizing

$$la\ +\ avoir\ mangée,$$
$$(\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell})\ (\mathbf{ip}_2^{\ell})\qquad (\mathbf{p}_2\ \mathbf{o}^{\ell}).$$

The past participle here should not be formed from the extended infinitive *la + manger* of type $\bar{\mathbf{i}}$, instead the extended infinitive of the auxiliary verb has type $\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\mathbf{p}_2^{\ell}$.

The verb *être* allows many kinds of complements, e.g. adjectives of type **a**, as in

$$\begin{array}{cc} \hat{e}tre & heureux \\ (\underbrace{\mathbf{ia}^{\ell}) & \mathbf{a}} \end{array}$$

in which case its past participle has type $\mathbf{p}_2\mathbf{a}^{\ell}$, as in

$$\begin{array}{ccc} avoir\ \acute{e}t\acute{e} & heureux \\ (\underbrace{\mathbf{ip}_2^{\ell}) & (\underbrace{\mathbf{p}}_2 & \underbrace{\mathbf{a}^{\ell})} & \mathbf{a}} \end{array}$$

It can also serve as an auxiliary verb to form the composite part of certain intransitive verbs as well as the passive of transitive verbs:

As far as we know, a past participle *été* of type $*\ \mathbf{p}_2\mathbf{p}_2'^{\ell}$ does not exist, since

$$\begin{array}{ccc} * \ avoir\ \acute{e}t\acute{e} & venu \\ (\underbrace{\mathbf{ip}_2^{\ell}) & (\underbrace{\mathbf{p}}_2 & \underbrace{\mathbf{p}_2'^{\ell})} & \mathbf{p}'_2} \end{array}$$

seems to be inadmissible. However, the following is allowed:

$$\begin{array}{cccc} avoir\ \acute{e}t\acute{e} & mang\acute{e} \\ (\underbrace{\mathbf{ip}_2^{\ell}) & (\underbrace{\mathbf{p}}_2 & \underbrace{\mathbf{o}^{\ell\ell}\mathbf{p}_2^{\ell})} & (\underbrace{\mathbf{p}}_2 & \underbrace{\mathbf{o}^{\ell})}} \end{array}$$

The types of past participles are covered by the following:

METARULE I. If the infinitive of the (non-extended) verb V has type $\mathbf{i}x^{\ell}$, then its past participle has type $\mathbf{p}_2x^{\ell}$ for most verbs, including all transitive verbs, and type $\mathbf{p}'_2x^{\ell}$ for a select group of intransitive verbs and for all reflexive verbs. The composite past of the former is formed with *avoir* of

type $\mathbf{ip}_2^{\ell}$, that of the latter with *être* of type $\mathbf{ip}_2'^{\ell}$.

We shall look at a few examples:

6. Finite verb forms.

To form a sentence we require the finite form of a verb. With any verb V in colloquial French there are associated $5 \times 6 = 30$ finite forms V_{jk}, where j ranges from 1 to 5 representing four simple tenses and the subjunctive mode:

present,	imperfect,	future,	conditional,	subjunctive,
1	2	3	4	5

and k ranges from 1 to 6 representing three persons singular followed by three persons plural. In literary French there are two additional tenses:

simple past, past subjunctive
6 7

For expository purposes, we shall ignore the last two, although they could be treated in the same way as the first five.

In this article, we shall assume the 30 finite forms as given, but the interested reader can look them up in [Bescherelle 1, 1998] or calculate them by the method of [Lambek 1976]. (Warning: the arrow there points in the opposite direction.)

Here, for example are the 30 finite forms of the verb *devoir*:

dois	*dois*	*doit*	*devons*	*devez*	*doivent*
devais	*devais*	*devait*	*devions*	*deviez*	*devaient*
devrai	*devras*	*devra*	*devrons*	*devrez*	*devront*
devrais	*devrais*	*devrait*	*devrions*	*devriez*	*devraient*

It is shown [loc.cit.] how these forms can be calculated from the following stems:

$$doi/s, \quad dev/ons, \quad doiv/ent, \quad dev/r/ai,$$

the stems appearing before the /. Similar calculations apply to all other verbs, the only exceptions being 10 frequently occurring verbs such as *aller, avoir, être, vouloir* and *pouvoir*.

What are the types of the finite forms? We assign the type $\mathbf{s}_j$ to a declarative sentence in the j-th tense ($j = 1$ to 4) and $\mathbf{s}_5$ to an incomplete subjunctive clause before the *que*. We assign the type π_k ($k = 1$ to 6) to the k-th personal subject pronoun:

$$\underset{\pi_1}{je}, \quad \underset{\pi_2}{tu}, \quad \underset{\pi_3}{il/elle/on}, \quad \underset{\pi_4}{nous}, \quad \underset{\pi_5}{vous}, \quad \underset{\pi_6}{ils/elles}.$$

We assign type $\overline{\pi}^r_k \mathbf{s}_j \bar{\mathbf{i}}^\ell$ to $(devoir)_{jk}$. The reason for the two bars will become clear later. For example, we have

$$\underset{\pi_5}{il} \quad \underset{(\overline{\pi}^r_3 \ \mathbf{s}_2 \ \bar{\mathbf{i}}^\ell)}{devait} \quad \underset{\mathbf{i}}{dormir},$$

a statement in the imperfect tense of type $\mathbf{s}_2$.

We expect to be able to type all finite forms of all verbs with extended infinitives of type $\mathbf{i}x^\ell$, $\bar{\mathbf{i}}x^\ell$ or $\bar{\bar{\mathbf{i}}}x^\ell$.

For example, we should be able to handle

dormir	of type	$\mathbf{i} = \mathbf{i}1 = \mathbf{i}1^\ell$	$(x = 1)$
manger	of type	$\mathbf{io}^\ell$	(x=**o**)
en + manger	of type	i	$(x = 1)$
la + manger	of type	i	$(x = 1)$
donner	of type	$\mathbf{i}\omega^\ell\mathbf{o}^\ell = \mathbf{i}(\mathbf{o}\omega)^\ell$	$(x = \mathbf{o}\omega)$
	or	$\mathbf{io}^\ell\omega^\ell = \mathbf{i}(\omega\mathbf{o})^\ell$	$(x = \omega\mathbf{o})$
lui + donner	of type	$\bar{\mathbf{i}}\bar{\mathbf{o}}^\ell$	$(x = \mathbf{o})$
la + donner	of type	$\bar{\mathbf{i}}\omega^\ell$	$(x = \omega)$
en + donner	of type	$\mathbf{i}\omega^\ell$	$(x = \omega)$
la + lui + donner	of type	$\bar{\mathbf{i}}$	$(x = 1)$
lui + en + donner	of type	$\bar{\mathbf{i}}$	$(x = 1)$

Direct sentences.

We can now state the following:

METARULE II. If the extended verb V has type $\mathbf{i}x^\ell\bar{\mathbf{i}}x^\ell$ or $\bar{\bar{\mathbf{i}}}\,x^\ell$, its finite form V_{jk} has type $\overline{\pi}^r_k\mathbf{s}_j x^\ell$ in a direct declarative sentence.

To extend the metarule to direct questions, we introduce the following basic types:

$\mathbf{q}_j$ for direct questions in the j-th tense ($j = 1$ to 4) subject to the following ordering:

$$\mathbf{q}_j \longrightarrow \mathbf{q} \longrightarrow \overline{\mathbf{q}}.$$

METARULE II (continued). $V_{jk}(j = 1\ mboxto\ 4)$ has type $\mathbf{q}_j x^\ell \pi^\ell_k$ in inverted direct questions.

Direct questions can also be formed without inversion from a declarative sentence by prefixing *est-ce que* of type $\mathbf{qs}^\ell$.

Both direct statements and inverted direct questions are fromed from the extended infinitive. Here are some examples:

From *en + lui + donner* of type **i** we form

$$\begin{array}{c} \textit{nous}\quad \textit{en}\ +\ \textit{lui}\ +\ \textit{donnons} \\ \pi_4 \qquad\qquad\quad (\pi^r_4\mathbf{s}_1) \end{array}$$

and

$$\begin{array}{c} \textit{en}\ +\ \textit{lui}\ +\ \textit{donnons} - \textit{nous?} \\ (\mathbf{q}_1\pi^\ell_4) \qquad\quad \pi_4 \end{array}$$

From *manger* of type $\mathbf{io}^\ell$ we form

$$\begin{array}{c} \textit{il mange}\ \underbrace{\textit{une pomme}} \\ \pi_3\ (\overline{\pi}^r_3\mathbf{s}_1\mathbf{o}^\ell)) \qquad \mathbf{o} \end{array}$$

and

$$\begin{array}{c} mange-t-il\ \underbrace{une\ pomme}?\\ (\mathbf{q}_1\mathbf{o}^\ell\pi_3^\ell)\quad\ \pi_3\end{array}$$

We must blame the Académie Française for not insisting that *mange* be spelled with a silent t at the end. We are told that

$$\begin{array}{c} mangé\text{-}je\quad\underbrace{une\ pomme}?\\ \underbrace{(\mathbf{q}_1\mathbf{o}^\ell\underbrace{\pi_1^\ell}_1)\quad\pi}_1\quad\ \mathbf{o}\end{array}$$

while acceptable, with the unusual spelling *é*, is better avoided.

We postulate $\mathbf{n}_s\to\overline{\pi}_3,\ \mathbf{n}_p\to\overline{\pi}_6$, so

$$\begin{array}{c} Jean\ mange\quad\underbrace{une\ pomme}_{\mathbf{o}},\\ \underbrace{\mathbf{n}_3}_3\quad\underbrace{(\overline{\pi}^\ell\mathbf{s}_1\mathbf{o}^\ell)}_{\mathbf{o}}\end{array}$$

and

$$\begin{array}{c}\underbrace{des\ professeurs}_{\pi_6}\ en\ +\ mangent\\ \underbrace{\qquad\qquad}_6\ (\overline{\pi}^r\mathbf{s_1})\end{array}$$

are acceptable. Finally, the bar on $\overline{\pi}_k$ guards against

$$\begin{array}{c}{}^*mange\ Jean\ \underbrace{une\ pomme}_{\mathbf{o}}\ ?\\ (\mathbf{q}_1\mathbf{o}^\ell\underbrace{\pi_3^\ell}_3)\quad\overline{\pi}_3\end{array}$$

and

$$\begin{array}{c}{}^*en\ +\ mangent\quad\underbrace{les\ professeurs}_{\pi_6}\ ?\\ (\mathbf{q}_1\underbrace{\pi_6^\ell}_6)\quad\overline{\pi}_6\end{array}$$

There are other ways of asking direct questions with the help of special question words such as *pourquoi* and *qui*. We shall assign the type $\overline{\mathbf{q}}$ to such questions, hence $\overline{\mathbf{q}}\mathbf{q}^\ell$ to *pourquoi*, as in

$$\begin{array}{c} pourquoi\quad vient-il?\\ \underbrace{(\overline{\mathbf{q}}\mathbf{q}^\ell)}_1\ \underbrace{(\mathbf{q}_1\pi_3^\ell)}_3\ \pi_3\end{array}$$

The bar is required to avoid

$$\begin{array}{c}{}^*pourquoi\ pourquoi\\ \underbrace{(\overline{\mathbf{q}}\mathbf{q}^\ell)\ (\overline{\mathbf{q}}\mathbf{q}^\ell)}\end{array}$$

The word *qui* can ask either for the subject or for the object. In the former case it has type $\overline{\mathbf{q}}\mathbf{s}^\ell\pi_3$, in the latter case type $\overline{\mathbf{q}}\mathbf{o}^{\ell\ell}\mathbf{q}^\ell$:

INSERT FORMULA HERE

These two occurrences of *qui* have distinct translations into German (*wer*/*wes*) or into pedantic English (*who*/*whom*).

INSERT SECTION 8 HERE

9. Some final words.

In this provisional attempt to describe French sentence structure by computations on types, we have necessarily confined ourselves to a small part of French grammar, and it goes without saying that some of our type assignments may have to be revised when further work is done. In particular, we have not yet looked at any but the most rudimentary noun-phrases and we have completely omitted from our investigation adverbs, relative clauses, negatives and imperatives. While many of these topics can be included in our framework, more serious problems may arise if we try to incorporate essential distinctions between masculine and feminine, between singular and plural and between persons and things. These distinctions often require semantic and pragmatic considerations outside our scope.

To look at only one example, in

$$\begin{array}{c} la \ + \ avoir \quad mangée \\ (\bar{\mathbf{i}}\mathbf{o}^{\ell\ell}\bar{\mathbf{i}}^{\ell})(\mathbf{i}\mathbf{p}_2^{\ell}) \qquad \mathbf{p}\_\mathbf{o}^{\ell} \\ \underbrace{\qquad\qquad\qquad\qquad}_{2} \end{array}$$

the silent e at the end of the past participle is not audible and could be ignored in analyzing spoken French. However, this excuse won't work with

$$la \ + \ avoir \ prise.$$

Conceivably, we could extend our treatment to account for this gender mark, but in

j'ai été heureuse

we cannot account for the ending of the adjective on syntactic grounds at all: we must know the sex of the speaker.

The mathematical analysis underlying our approach was first explored in [Lambek 1999] and its history was also discussed in [Casadio and Lambek, to appear].

11. Response to referee's comments.

Not everybody will be happy about our proposal. Referee 1 objects to our attempt to lump syntactic categories, morphological features and grammatical functions under the single heading of what we call "types". On the other hand, he or she criticizes us for not incorporating semantics.

Of course, *lexical* semantics has to be stored in the dictionary; but *functional* semantics, as in Montagne grammar, could in principle be derived from the structure of compound types. For example, ab^{ℓ} and $b^{r}a$ could be interpreted as denoting functions from the set of entities of type b to the set of entities of type a. However, to fully justify such an interpretation, one should adopt a more elaborate algebraic system, namely one derived

from "classical bilinear logic", as proposed by Claudia Casadio, but at the cost of making computations more difficult. For a more thorough discussion of this question, see the article "A tale of four grammars", to appear in Studia Logica.

The referee also raises the question of how to block such sentences as "*to eat an apple on one foot sleeps*". This could be blocked by suitable type assignments if one took the trouble; but it seems more reasonable to regard it as acceptable *syntactically*. Indeed, attempts to block it would also block "*to eat an apple on one foot makes one's foot fall asleep*" and even Chomsky's "*colourless green ideas sleep furiously*".

Referee 2 wants to know the limitations of our approach and whether it also applies to other languages. The approach has been applied to English, German and Italian. Admittedly, these are all Indo-European languages, but first steps are being taken to look at some non-Indo-European languages as well, e.g. Arabic.

There are of course serious limitations to our approach, even for English. For example, if the word "*whom*" is omitted in "*people (whom) John likes like him*", there is no word left to which the type of "*whom*" can be attached. Indeed, it becomes necessary to admit some grammatical rules other than those encoded in the types stored in the dictionary.

Referee 3 wants to know whether we can predict the correct order of preverbal clitic pronouns, which is well-known to all teachers of French. This is precisely what we have been trying to do, by carefully choosing appropriate type assignments and by fine-tuning the partial order in the set of basic types.

REFERENCES

Bescherelle 1, *L'Art de conjuger*, Éditions Hurtubise, Montréal 1998.

Bescherelle 3, *La grammaire pour tous*, Éditions Hurtubise, Montréal 1998.

J.-P. Boons, A. Guillet and C. Leclerc. *La structure des phrases simples en français, Constructions intransitives*, Librairie Droz, Genève-Paris 1976.

C. Casadio and J. Lambek. A tale of four grammars, *Studia Logica* (to appear).

M. Gross. *Grammaire transformationelle du français, Syntaxe du verbe*, Librairie Larousse, Paris 1968.

M. Gross. *Table des verbes entrant dans des constructions complètives*, Éditions du CNRS, Paris 1969.

A. Guillet and C. Leclerc. *La structure des phrases simples en français, Constructions transitives locatives*, Librairie Droz, Genève Paris 1992.

Z. Harris, A cycling cancellation-automaton for sentence well-formedness, *International Computation Center Bulletin* 5 (1966), 69-94.

R. Jackendoff, $\overline{X}$ *Syntax: A Study of Phrase Structure*, The MIT Press, Cambridge, Mass., 1977.

J. Lambek, A mathematician looks at French conjugation, *Linguistic Analysis* 2 (1976), 203-214.

J. Lambek, Type grammar revisited, in: A. Lecomte, F. Lamarche and G. Perrier (eds). *Logical aspects of computational linguistics*, Springer LNAI 1582(1999), 1-27.

FOOTNOTE:

The authors acknowledge support from the Social Sciences and Humanities Research Council of Canada.