A Declarative Approach to Onset Maximisation

Alain Thériault
Département de linguistique et de traduction
Université de Montréal

Abstract

A consonantal cluster is sometimes structurally ambiguous according to the phonotactic grammar of French. Actually, a maximum of consonants must be associated to the onset of the right-hand syllable. '*Marbré*' is syllabified as /mar·bre/, not */marb·re/. In Declarative Phonology, Onset Maximisation must be realised by means of unification, not by a resyllabification process that modifies a pre-specified structure. In this paper, I propose a declarative analysis of onset maximisation in the form of a constraint on syllable strings.

Résumé

Un groupe consonantique est parfois structuralement ambigü selon la grammaire phonotactique du français. En fait, un maximum de consonnes doit être associé à l'attaque de la syllabe de droite. '*Marbré*' est syllabifié /mar·bre/, non */marb·re/. En phonologie déclarative il faut réaliser la 'Maximisation des attaques' par unification, et non pas par une 'resyllabification', modifiant une structure préalablement spécifiée. Je propose ici une analyse déclarative de la maximisation des attaques, qui prend la forme d'une contrainte sur l'enchaînement des syllabes.

1. Introduction

Starting from the assumption that syllable structure is predictable, a system allowing a string of segments to be associated to this structure is necessary. Many systems have been proposed during the last twenty five years. One of the problems these theories faced was the division of consonantal clusters. Some authors, (Kahn 1976, Stériade 1982, Clements et Keyser 1983, Levin 1985 to name just a few) proposed a rule approach to syllabification. The second rule stipulates that intervocalic clusters are to be divided according to the Onset Maximisation Principle. This principle will associate a maximum of segments to the onset of the following syllable. This association must respect the language phonotactics. As simple as this solution may seem, it is not so simple to explain to a computer.

Since January 1995, Professor John Reighard of Université de Montréal, my colleague Jocelyn Gagnon and I have been working on a parser that associates a syllable structure to a string of characters representing the phonemes of a lexical entry. This parser is described within the Declarative Phonology (DP) framework, as proposed by Scobbie (1991 and 1993). DP is based on a formalism already in use in a family of syntactic and semantic models (GPSG, HPSG, LFG, FUG and others). The formalism is based on attribute-value structures, and is used in a framework in which the only operation allowed is unification. This declarative

framework is thus a challenge to phonology where objects have been traditionally described procedurally, by specifying a construction and/or destruction algorithm 'transforming' a phonological representation into a phonetic representation.

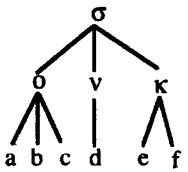
This paper is divided as follows:

- description of objects (the representation of information)
- the definition and description of the theoretical framework
- associated problems of onset maximisation
  - the Reighard-Gagnon-Thériault parser
  - the solution I propose in the DP framework
- conclusion

2. The description of objects (the representation of information)

I will assume here that the syllable is a three level multibranching structure in which each branch can be multibranching. The maximum number of branches allowed is prescribed by the phonotactics of the language described. In French, a maximal syllable is composed of a three branch onset, a single branch nucleus and a two branch coda (fig. 1)

fig. 1[1]



As to French phonotactics, a description could be made in terms of re-write rules. This 'phonotactic grammar' (1) with its phonotactic constrainst (2) is largely inspired by the one Yves-Charles Morin (personal communication) proposed.[2]

(1) σ -> (o) v (κ)
o-> ([-son, +cont, +cor]) C ([+son, -nas])
v -> V
κ -> (r) C[3]
(C = any [- voc] segment)

---

[1] σ= syllable, o= onset, v= nucleus and κ= coda. "a,b...f" represent segments.

[2] Morin's grammar includes a compulsory "rime" tier containing the nucleus and the optional coda. He also provides for a double nucleus in case of 'light' diphthongs such as the [wa] in *trois* [trwa]

[3] Morin's grammar allows an optional "s" right of "C". In Thériault (1996) I argue for a strategy of consistent onset maximisation that assigns a different structure to "s" in that position.

(2) A few phonotactic constraints[4]

1. If [CL] $_{O}$, then C = [-cont] or [-son, -cor]

(i.e. if a liquid is the second of a two member onset cluster then the first is a stop or a labial fricative)

2. If [Cl] $_{O}$, then C ≠ [-son, -cont, +cor]

(i.e. if the second of a two member onset cluster is /l/ then the first is non coronal: that is, *tl-, *dl- are ungrammatical onsets)

3. The theoretical framework

3.1. DP (Declarative Phonology)

Since the early 1990's, many scholars such as James M. Scobbie, Steven Bird or T. Marc Ellison have proposed approaching phonology with unification-based theory[5]. In the declarative approach, a phonological object (phoneme, syllable, etc.) is a set of constraints. These constraints are to be expressed in a declarative language (non-procedural), they are non-destructive and non-conflicting. The phonological information is expressed in terms of partial information structures (also known as feature structures or attribute-value structures) on which unification is the only admissible operation. Since DP is monostratal, there are no derivations, nor does it allow for a level where a phonotactically incorrect form could arise, awaiting repairs as in the Constraints and Repair strategies theory (Paradis 1988, Singh 1990).

Since DP is a framework, it needs a theoretical base on which to work. The theoretical base used here is the theory of Syllabic Templates (Selkirk, 1982). I also use the Rule Approach to Syllabification (Kahn 1976, Stériade 1982, Clements and Keyser 1983, Levin 1985), to describe the syllable structure.

3.2 The Syllabic Template theory

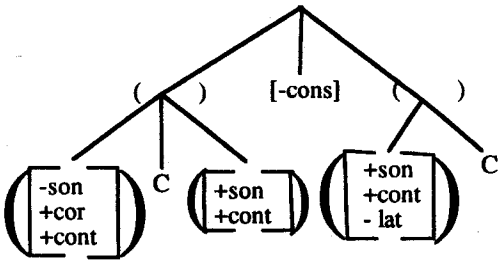
In the Syllabic Template theory, all nodes under the SYLL (syllable) node must contain only permissible elements according to the phonotactic (1) grammar. This theory provides a test to determine whether the distribution of segments corresponds to the phonotactics of the language. If the template test fails, then the string cannot be syllabified. The phonotactic grammar (1) of French gives the following template (fig. 2).

---

[4] These are the most obvious constraints. It is clear that more are needed, such as the ones that would describe the environment of the nasal consonants.

[5] This approach was named Constraint-based Phonology by Bird (1990) while Scobbie (1991) used the expression Declarative Phonology.

fig. 2



3.3. The Rule Approach to Syllabification

The Rule Approach to Syllabification can be described as in (3) for many languages such as French[6].

(3) (i) associate all vocalic segments to the NUC (nucleus) position
(ii) associate a maximum of non-vocalic segments preceding the nucleus to the ONS (onset) position of the same syllable
(iii) associate all remaining segments to the CODA (coda) position of the preceding syllable.

The main advantage of the rule approach is that it introduces the notion of onset maximisation (3 ii).

4. Onset maximisation

Onset maximisation assigns a maximum number of segments to the onset of the following syllable (even if this syllable doesn't have a nucleus)[7]. A strict application of the maximal onset principle removes any structural ambiguity in consonantal clusters. For example, a word like *marbré*, Eng. 'marbled', has an ambiguous cluster /rbr/ since it can be syllabified in two different ways (4) according to the phonotactic grammar.

---

[6]The details of this description do not necessarily apply to all languages. English, for example, allows non-vocalic segments in certain nuclei.

[7]What is traditionally called extrasyllabic segments, in French at least, appear only in word initial or word final position, and conform partially to normal rules of syllabification. Since initial and final appendices are structurally and distributionally indentical to a coda (initial appendix) and to an onset (final appendix) I argue in Thériault (1996) that those segments are in fact part of "nucleus-free" syllables that are allowed only at the begining or at the end of a word. A word is thus a set of at least one nuclear syllable (a syllable with a nucleus) that can be preceded and/or followed by a nucleus-free syllable.

(4) a /mar·bre/[8]
b /marb·re/

The maximal onset principle selects (4 a) as the correct form since a maximum of segments must be associated to the onset of the following syllable. Onset maximisation also prevents an onset-free syllable from being preceded by a syllable in which a segment is associated to the coda. For example the word *manie*, Eng. 'fad', can also be syllabified in two different ways (5).

(5) a. /ma·ni/
b. /man·i/

Again, both forms are accepted by the phonotactic grammar although only (5 a) is acceptable in French.

It may seem easy to apply onset maximisation, since all one needs to know is the phonotactic grammar. From a declarative point of view however, it is not so simple. Before introducing the difficulties related to cluster division in DP, I will succinctly present the Reighard-Gagnon-Thériault parser.

4.1 The Reighard-Gagnon-Thériault parser

The parser we are currently developing allows a string of characters, representing a lexical entry's phonemes, to be associated to a syllabic structure and to give a phonetic representation corresponding to Montréal French. In addition to the syllabifier the parser contains different modules that give surface representations containing the phonological characteristics of French (vowel lengthening and tension, accent setting, diphtongisation, etc).

The parser is built within a Prolog program in which we apply the classical technique of differential lists associating strings of phonemes (the lexical entries) to syntagmatic structures (the syllabic structure). The latter are defined by the phonotactic grammar (1).

The first module we developed was the syllabifier since many phonotactic constraints are associated with properties of the syllable. A syllable is defined as a triplet (6):

(6) SYLL(ONS(o.o'.o''.r),NUC(n.r'),CODA(k.k'.r''))

To this structure are applied constraints specifying the segmental structure of each of the constituents, exactly as provided for by the phonotactic grammar (1). A lexical entry is represented as a list of phonemes, and the differential list technique amounts in fact to a left-to-right scan, associating sublists of the list of phonemes to legal syllable positions. However, since the legal syllable positions are defined only in terms of individual syllables, the parser returns all possible grammatical associations, including, for lexical entries such as [marbre] and [mani], (4 a and b) and (5 a and b).

---

[8] The '·' identifies the syllable cut.

From the point of view of the structure of individual syllables, the system is in effect restricted to those syllable structures defined by the phonotactic grammar. One of its interesting features is therefore that it can distinguish between grammatical and ungrammatical syllables. What is needed in addition, obviously, is the maximal onset constraint, in order to rule out (4 b) and (5 b).

In Declarative Phonology, onset maximisation has to be realised by unification. A "re-syllabification" that modifies a previously specified structure is unacceptable. Phonemes have to be unified to a structure.

4.2 A solution

Onset maximisation prevents two things: (i) an empty onset preceded by a coda containing one (or two) element(s) such as (5 b) and (ii) an onset preceded by a coda in which the last element could have been parsed in the following onset such as (4 b). These two conditions can be summarised into one condition (7).

(7) The last element of a coda can not be an admissible element of the following onset.

This condition, for a cluster composed of the string x,y,z, can be expressed in form of logic as in (8).

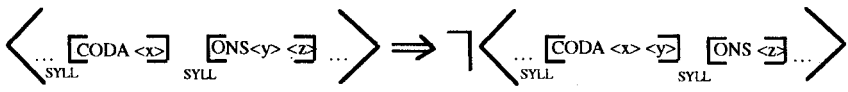
(8)

< ... CODA(x) + ONS(y,z) ... > -> ¬< ... CODA(x,y) + ONS(z) ... >
(where '+' indicates linear precedence)

This condition prevents a cluster from being syllabified twice. If a string of segments, containing the segments x, y, z where x and z are optional and z is not necessarily the last element of the cluster, can be syllabified as follows: $_{SYLL}$[CODA(x)] $_{SYLL}$[ONS(y,z)] then it can not be syllabified: $_{SYLL}$[CODA(x,y)] $_{SYLL}$[ONS(z)].

In a declarative analysis, onset maximisation takes the form of a constraint on a string of syllables, not on the syllables. The condition presented in (8) applies to a list of syllables and can be expressed in a DP formalism as follows (fig. 3).

fig. 3



This constraint on syllable strings within a word applies between every pair of elements within a string. Every pair of syllables has to correspond to the phonotactic grammar and to the constraint. That is, for a list of syllables a.b.c.d.nil. each pair [a.b], [b.c] and [c.d] has to obey to the constraint. This implies the identification of every pair of syllables and of every last element of the first syllable of the pair. The construction of an admissible onset demands that this last element is to be incorporated to the already existing material included in the onset.

5. Conclusion

This approach to onset maximisation allows for an interesting treatment of traditionally called "extrasyllabic" material, which in French is limited to word initial and word final appendices. By applying onset maximisation to all consonant clusters, I can eliminate distinctions such as Word-Initial-Onset vs. Word-Internal-Onset, and Word-Final-Coda vs. Word-Internal-Coda[9]. The elimination of such distinctions allows a simplification of the syllable description since syllable positions are attributed in a consistent manner. Since a word initial appendix in French is a coda, and a word final appendix is an onset, only two kinds of syllables need to be defined, those that have a nucleus and those that do not. This distribution is ensured by another constraint on syllable strings, limiting non nuclear syllables to initial and final positions in the string of syllables.

The formalism associated to a phonology that allows only the addition of information and disallows destructive procedure or the manipulation of phonological forms (rule ordering or constraint ranking) is still quite recent but it opens the door to a more transparent description by the use of a metalanguage independent of the described objects.

Acknowledgements

This paper is a modified version of the presentation I gave at the Student Conference in Computational Linguistics in Montreal in June 1996. This English version was presented at the Holland Institute of Generative Linguistics (HIL) at Leiden University in July 1996. I would like to thank Prof. John Reighard for his constant assistance and suggestions. Of course, I remain responsible for any errors, theoretical or formal, that could have slipped into this paper.

[9]Such distinctions can be found in Bird and Klein (1994: 31).

REFERENCES

Bird, Steven (1990), *Constraint-based phonology*, PhD Dissertation, University of Edinburgh.

---------- & Ewan Klein (1994), *Enriching HPSG phonology*, Centre for Cognitive Science, University of Edinburgh.

Clements, George N. & Samuel J. Keyser (1983), *CV phonology: a generative theory of the syllable*, Linguistic Inquiry Monograph 9, MIT Press, Cambridge, Mass.

Kahn, Daniel, M. (1976), *Syllable based generalization in English phonology*, PhD Dissertation, MIT, Cambridge, Mass.

Levin, Juliette, M. (1985), *A metrical theory of syllabicity*, PhD Dissertation, MIT, Cambridge, Mass.

Paradis, Carole (1988), "On constraints and repair strategies" *The linguistic review* 6, pp. 71-97.

Scobbie, James M. (1991), "Towards Declarative Phonology", *Edinburgh Working Papers in Cognitive Science* 7, pp 1-27.

---------- (1993), "Constraint violation and conflict from the perspective of declarative phonology", dans Paradis, Carole & Darlene Lacharité (eds). *Constraint based theories in mulitlinear phonology*, CJL/RCL Special Issue. 38(2), pp. 155 167.

Selkirk, Elisabeth O (1982), "The syllable", Hulst, Harry van der & Norval Smith (eds) *The structure of phonological representaions* Part 2, Floris, Dordrecht, pp. 336-383.

Singh, Rajendra (1990), "Vers une théorie phonotactique générative", *Revue québécoise de linguistique* 19(1), pp. 131-163.

Steriade, Donca, M. (1982), *Greek prosodies and the nature of syllabification*, PhD Dissertation, MIT., Cambridge, Mass.

Thériault, Alain (1996), *Description de la syllabe française; une approche déclarative*, Masters Thesis, Université de Montréal.

Alain Thériault
Université de Montréal
Département de linguistique et de traduction
C.P. 6128 Succursale Centre-ville
Montréal (Qc) H3C 3J7

E-mail: theriaal@ere.umontreal.ca
Fax: (514) 343-2284