

Colloque "Le patrimoine à l'ère du numérique"
10 et 11 décembre 2009
Université de Caen Basse-Normandie

Éliane Delente & Richard Renault
CRISCO, EA 4255
Université de Caen

Annotation automatique de textes versifiés

- 1- présentation du projet ANAMÈTRE
- 2- les étapes du traitement automatique

1- présentation du projet ANAMÈTRE

Le projet Anamètre

- des linguistes et des métriciens des universités de Nantes, Paris 8 et Caen (Crisco)
- des informaticiens (Greyc, Caen)

Objectifs

Trois volets :

- 1- la constitution d'une base de données de textes poétiques et théâtraux annotés, du XVIIe au début du XXe siècle
- 2- la constitution d'une base de données de relevés métriques
- 3- l'élaboration d'outils d'analyse métrique automatique : calcul du formatage métrique des vers et du formatage strophique des poèmes

L'état de l'art

Frantext, Gallica, Enclitt : des textes impropres à une exploitation métrique

Des relevés métriques en format propriétaire et/ou obsolète

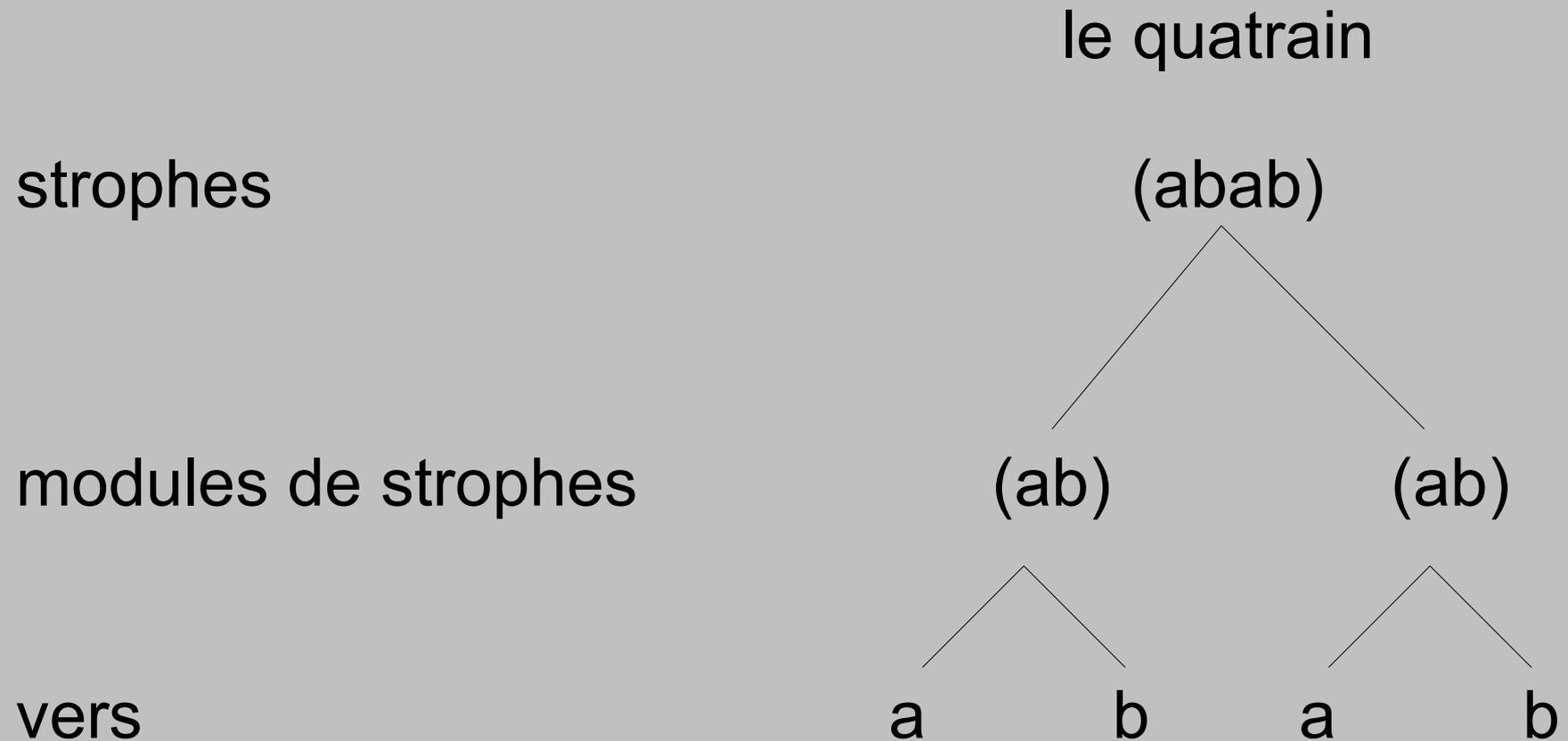
Un seul outil de traitement automatique de la métrique (Beaudouin, 2002) mais limité au théâtre classique, non disponible et limité dans le calcul des structures strophiques

Pourtant...

Les textes métriques : un matériau idéal pour l'étude de corpus et le traitement automatique

- 1- périodicité : les formes métriques se répètent à l'échelle du texte
- 2- systématité : un poème est généralement entièrement métrique ou non métrique

Les regroupements métriques : les strophes



Les regroupements métriques : les vers

l'alexandrin

vers

(6-6)

hémistiches

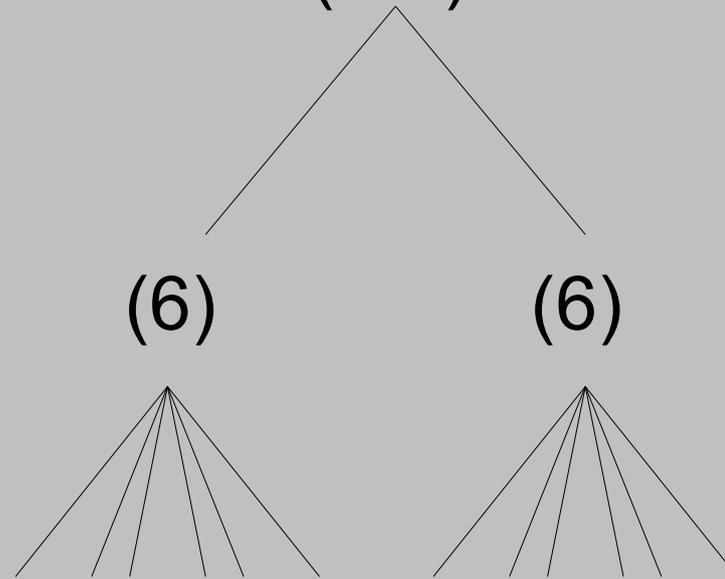
(6)

(6)

voyelles métriques

1 2 3 4 5 6

1 2 3 4 5 6



Qu'est-ce que la métrique ?

L'art de toujours retomber sur ses pieds

4

6

sous-vers 1

sous-vers 2

rime

= 34 graphèmes

L'énoncé ci-dessus n'est pas un vers. Hors contexte, c'est un morceau de prose.

La constitution d'une base de données de textes versifiés

Examen critique des sources éditoriales

La correction de textes déjà numérisés

La ponctuation

L'orthographe : *temps* ne rime pas avec *sang*

La numérisation

Le format : XML et standard de la TEI

La constitution du corpus

Trois paramètres plus un :

- 1- La période : du début XVIIe au début du XXe siècle
 - 2- Le répertoire des mètres français : simples et composés
 - 3- Les formes strophiques : simples et composées
- ... et les formes fixes échappant à la périodicité : sonnet, ballade, rondeau, triolet etc.

L'annotation : unités inférieures au vers

Les e féminins

Les semi-consonnes (diérèses)

Les voyelles métriques

Les sous-vers

La tonique de sous-vers et de vers

La rime (qui comprend la dernière voyelle stable et les phonèmes qui la suivent : *belles* / *éternelles*)

La PGTC (la plus grande terminaison commune comprend ici, en plus de la rime, la consonne d'attaque de la dernière voyelle stable : *attitudes* / *études*)

L'annotation

Numérotation : la voyelle métrique, le sous-vers, le vers, le module de strophe et la strophe

La transcription phonétique : voyelles métriques et mot à la rime

Le formatage métrique du vers. Ex. : 8, 4-6, 6-6 etc.

Le schéma de rimes : abba etc.

La forme globale du poème

La constitution d'une base de données de relevés métriques

Qu'est-ce qu'un relevé métrique ? Une série d'analyses métriques de pièces métriquement cohérentes et autonomes.

Pourquoi faire des relevés métriques ?

- offrir une vue synthétique de la métrique d'un auteur ou d'une époque
- faciliter la consultation et la comparaison des données massives
- permettre de repérer et de décrire des évolutions dans la pratique métrique
- permettre d'appréhender la métrique dans une perspective historique

Les données : de la main à l'outil.

Objectifs atteints :

- récupération des données
- pérennisation en utilisant un format ouvert

En cours...

- intégration dans une base de données
- normalisation et optimisation
- accessibilité à l'exploitation pour tous

L'élaboration d'outils d'analyse métrique automatique

Trois programmes de calcul automatique

- calcul du formatage métrique des vers
- calcul du formatage strophique
- calcul de la forme globale du poème

Le calcul du formatage métrique

La perception de l'égalité métrique

Quelles sont les unités métriques pertinentes ?

- l'égalité en nombre de consonnes (3) n'est pas perceptible : *plante, calter, trouve, oubliait, jambage* etc.
- l'égalité en nombre de voyelles (3) est immédiatement perceptible : *tombera, nuitamment, composé, opéra, employaient, aéré* etc.

Le calcul du formatage métrique

Quatre modules préalables :

- l'identification des voyelles
- l'e traitement de l'e instable
- le traitement des semi-consonnes
- le repérage de certaines unités prosodiques et morphosyntaxiques

Calcul du formatage strophique

- repérage des rimes
- repérage de leur agencement
- formulation d'hypothèses sur le formatage strophique

Calcul de la forme globale du poème

- les formes périodiques
- les vers-préludes
- les vers-clausules
- les formes fixes

Pourquoi ces outils ?

- libérer le chercheur de certaines tâches manuelles
- dynamiser la recherche en métrique
- un nouvel enjeu pour la stylométrie et la lexicométrie
- visées didactiques

Dynamiser la recherche en métrique

- tester les hypothèses avancées dans les travaux récents
- développer les études sur la concordance : interaction entre l'organisation syntaxique-sémantique du texte poétique et sa structuration métrique : strophes, modules de strophes, vers et sous-vers

Un nouvel enjeu pour la stylométrie et la lexicométrie.

de nouvelles unités balisées : à la strophe et au vers s'ajoutent le module de strophe, le sous-vers, la rime et chaque voyelle métrique caractérisée par sa nature prosodique et son statut rythmique

une nouvelle possibilité pour des analyses plus fines et surtout plus pertinentes sur l'articulation sémantique / rythme

Visées didactiques

- approche de la langue des vers (élision, diérèse)
- approche métrique
- approche stylistique : étude distributionnelle

2- les étapes du traitement automatique

- **Le traitement métrique**
- **Le traitement XML**
- **Les unités pertinentes**

illustration à partir d'un vers de Victor Hugo : *Paroles de géant, XVIII, La légende des siècles*

Que Mars bête et sanglant, que Diane bourrue,

Première étape : identification des noyaux syllabiques

- repérage des segments graphiques comptant comme voyelles

a : voyelle stable
e : voyelle instable
i : semi-consonne
e : segment ambigu

Que Mars bête et sanglant, que Diane bourrue,

Première étape : identification des noyaux syllabiques

- repérage des segments graphiques comptant comme voyelles

a : voyelle stable
e : voyelle instable
i : semi-consonne
e : segment ambigu

Allez voir Hermione ; et content de lui plaire,

Racine, *Andromaque*, vers 665

Que Mars bête et sanglant, que Diane bourrue,

Première étape : identification des noyaux syllabiques

- repérage des segments graphiques comptant comme voyelles

a : voyelle stable
e : voyelle instable
i : semi-consonne
e : segment ambigu

Que Mars bête et sanglant, que Diane bourrue,

Première étape : identification des noyaux syllabiques

- repérage des segments graphiques comptant comme voyelles
- attribution d'une valeur numérique pour le décompte des voyelles

Que Mars bête et sanglant, que Diane bourrue,
 1 1 1 1 1 1 1 0 1 1 1 1 1

Première étape : identification des noyaux syllabiques

- repérage des segments graphiques comptant comme voyelles
- attribution d'une valeur numérique pour le décompte des voyelles
- attribution du symbole phonétique correspondant au graphème

Que Mars bête et sanglant, que Diane bourrue,

1	1	1	1	1	1	1	0	1	1	1	1	1
ə	a	ɛ	ə	e	ã	ã	ə	j	a	ə	u	y ə

Première étape : identification des noyaux syllabiques

- repérage des segments graphiques comptant comme voyelles
- attribution d'une valeur numérique pour le décompte des voyelles
- attribution du symbole phonétique correspondant au graphème

Deuxième étape : traitement des "e" instables

- modification de la valeur des "e" élidables

Que Mars bête et sanglant, que Diane bourrue,
 1 1 1 0 1 1 1 1 0 1 1 1 1
 ə a ε ə e ã ã ə j a ə u y ə

Première étape : identification des noyaux syllabiques

- repérage des segments graphiques comptant comme voyelles
- attribution d'une valeur numérique pour le décompte des voyelles
- attribution du symbole phonétique correspondant au graphème

Deuxième étape : traitement des "e" instables

- modification de la valeur des "e" élidables
- modification de la valeur des "e" surnuméraires

Que Mars bête et sanglant, que Diane bourrue,
 1 1 1 0 1 1 1 1 0 1 1 1 1 1
 ə a ε ə e ã ã ə j a ə u y ə

Première étape : identification des noyaux syllabiques

- repérage des segments graphiques comptant comme voyelles
- attribution d'une valeur numérique pour le décompte des voyelles
- attribution du symbole phonétique correspondant au graphème

Deuxième étape : traitement des "e" instables

- modification de la valeur des "e" élidables
- modification de la valeur des "e" surnuméraires

Troisième étape : traitement des diérèses

- remplacement d'une semi-consonne par la voyelle correspondante si le mot appartient au dictionnaire des diérèses

Que Mars bête et sanglant, que Diane bourrue,
 1 1 1 0 1 1 1 1 1 1 1 1 1 1
 ə a ε ə e ã ã ə i a ə u y ə

Première étape : identification des noyaux syllabiques

- repérage des segments graphiques comptant comme voyelles
- attribution d'une valeur numérique pour le décompte des voyelles
- attribution du symbole phonétique correspondant au graphème

Deuxième étape : traitement des "e" instables

- modification de la valeur des "e" élidables
- modification de la valeur des "e" surnuméraires

Troisième étape : traitement des dièses

- remplacement d'une semi-consonne par la voyelle correspondante si le mot appartient au dictionnaire des dièses

Quatrième étape : décompte des voyelles

- numérotation des voyelles

Que	Mars	bête	et	sanglant,	que	Diane	bourrue,						
1	1	1	0	1	1	1	1	1	1	1	1		
ə	a	ɛ	ə	e	ã	ã	ə	i	a	ə	u	y	ə
1	2	3		4	5	6	7	8	9	10	11	12	13

Première étape : identification des noyaux syllabiques

- repérage des segments graphiques comptant comme voyelles
- attribution d'une valeur numérique pour le décompte des voyelles
- attribution du symbole phonétique correspondant au graphème

Deuxième étape : traitement des "e" instables

- modification de la valeur des "e" élidables
- modification de la valeur des "e" surnuméraires

Troisième étape : traitement des diérèses

- remplacement d'une semi-consonne par la voyelle correspondante si le mot appartient au dictionnaire des diérèses

Quatrième étape : décompte des voyelles

- numérotation des voyelles
- calcul du nombre de voyelles
- calcul du nombre de voyelles métriques du vers

Que Mars bête et sanglant, que Diane bourrue,

1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
ə	a	ɛ	ə	e	ã	ã	ə	i	a	ə	u	y	ə	
1	2	3		4	5	6	7	8	9	10	11	12	13	12

Chaîne de traitement

fichier XML(TEI)

```

<?xml lang="fr" version="1.0" encoding="UTF-8" ?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" type="text" ?>
  <text ?>
    <p ?>
      <b ?>Que Mars bête et sanglant, que Diane bourrue,</b>
    </p>
  </text>
</TEI>

```

programmes

fichier XML(TEI)

```

<?xml lang="fr" version="1.0" encoding="UTF-8" ?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" type="text" ?>
  <text ?>
    <p ?>
      <b ?>Que Mars bête et sanglant, que Diane bourrue,</b>
    </p>
  </text>
</TEI>

```

transformation

<1>Que Mars bête et sanglant, que Diane bourrue,</1>

fichier XHTML

```

<?xml version="1.0" encoding="UTF-8" ?>
<html ?>
  <body ?>
    <b ?>Que Mars bête et sanglant, que Diane bourrue,</b>
  </body>
</html>

```

Chaîne de traitement

fichier XML(TEI)

```

<?xml lang="fr" version="1.0">
<doc>
  <title>
    <h1>Mars bête et sanglant, que Diane bourrue,</h1>
  </title>
  <text>
    <p>Mars bête et sanglant, que Diane bourrue,
    </p>
  </text>
</doc>

```

programmes

fichier XML(TEI)

```

<?xml lang="fr" version="1.0">
<doc>
  <title>
    <h1>Mars bête et sanglant, que Diane bourrue,</h1>
  </title>
  <text>
    <p>Mars bête et sanglant, que Diane bourrue,
    </p>
  </text>
</doc>

```

transformation

<1>Que Mars bête et sanglant, que Diane bourrue,</1>

fichier XHTML

```

<?xml version="1.0" encoding="UTF-8" standalone="no" type="text/html">
<html>
  <head>
    <title>Mars bête et sanglant, que Diane bourrue</title>
  </head>
  <body>
    <p>Mars bête et sanglant, que Diane bourrue,</p>
  </body>
</html>

```

<seg type="vs" rule="403" phoneme="a" value="1">a</seg>

Chaîne de traitement

fichier XML(TEI)

```

<?xml lang="fr" version="1.0">
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <fileDesc>
    <title>Mars bête et sanglant, que Diane bourrue,</title>
    <author>Mars bête et sanglant, que Diane bourrue,</author>
    <editor></editor>
    <pubInfo>
      <date></date>
      <place></place>
      <publisher></publisher>
    </pubInfo>
    <text>
      <p></p>
    </text>
  </fileDesc>
  <encodingDesc>
    <charmap></charmap>
    <conformTo></conformTo>
  </encodingDesc>
  <text>
    <p></p>
  </text>
</TEI>

```

programmes

fichier XML(TEI)

```

<?xml lang="fr" version="1.0">
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <fileDesc>
    <title>Mars bête et sanglant, que Diane bourrue,</title>
    <author>Mars bête et sanglant, que Diane bourrue,</author>
    <editor></editor>
    <pubInfo>
      <date></date>
      <place></place>
      <publisher></publisher>
    </pubInfo>
    <text>
      <p></p>
    </text>
  </fileDesc>
  <encodingDesc>
    <charmap></charmap>
    <conformTo></conformTo>
  </encodingDesc>
  <text>
    <p></p>
  </text>
</TEI>

```

transformation

fichier XHTML

```

<?xml lang="fr" version="1.0" encoding="UTF-8" standalone="no">
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <fileDesc>
    <title>Mars bête et sanglant, que Diane bourrue,</title>
    <author>Mars bête et sanglant, que Diane bourrue,</author>
    <editor></editor>
    <pubInfo>
      <date></date>
      <place></place>
      <publisher></publisher>
    </pubInfo>
    <text>
      <p></p>
    </text>
  </fileDesc>
  <encodingDesc>
    <charmap></charmap>
    <conformTo></conformTo>
  </encodingDesc>
  <text>
    <p></p>
  </text>
</TEI>

```

Que Mars bête et sanglant, que Diane bourrue,

type = vs
règle = 403
phonème = a
valeur = 1

Les différentes unités du calcul métrique (attribut de la balise <seg>)

1 - identification des voyelles et semi-consonnes

type="vs"
voyelle stable

Mars

type="vi"
voyelle instable

bête

type="sc"
semi-consonne

Diane

Les différentes unités du calcul métrique (attribut de la balise <seg>)

1 - identification des voyelles et semi-consonnes

type="vs"
voyelle stable

type="vi"
voyelle instable

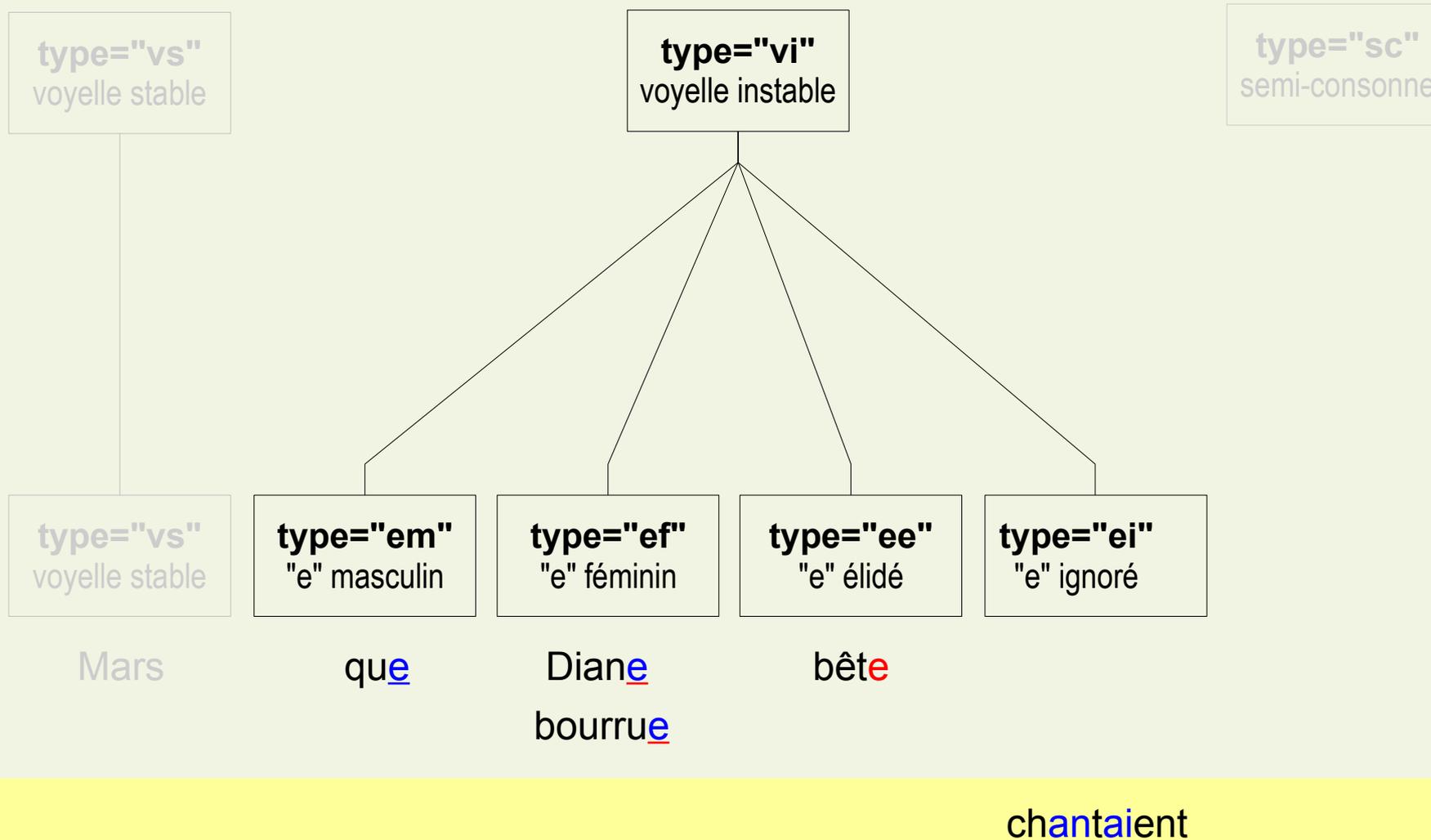
type="sc"
semi-consonne

type="vs"
voyelle stable

Mars

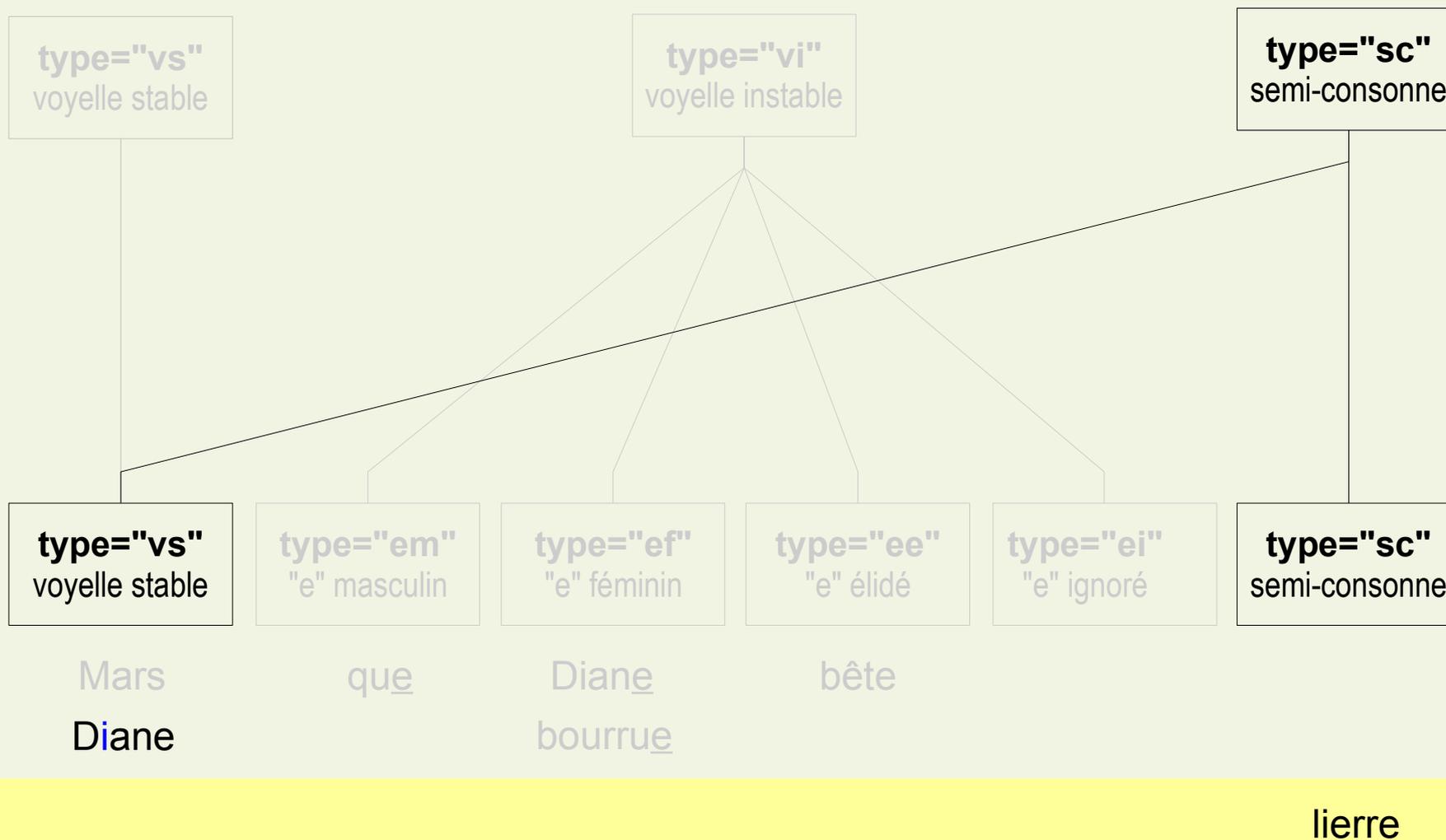
Les différentes unités du calcul métrique (attribut de la balise <seg>)

2 – traitement des "e" instables



Les différentes unités du calcul métrique (attribut de la balise <seg>)

3 – traitement des diérèses



Les différentes unités du calcul métrique (attribut de la balise <seg>)

unités pertinentes pour le calcul du nombre de voyelles (13)

type="vs"
voyelle stable

type="em"
"e" masculin

type="ef"
"e" féminin

type="ee"
"e" éliidé

type="ei"
"e" ignoré

type="sc"
semi-consonne

Que Mars bête et sanglant, que Diane bourrue,

em vs vs vs vs vs em vs vs ef vs vs ef

1 2 3 4 5 6 7 8 9 10 11 12 13

Les différentes unités du calcul métrique (attribut de la balise <seg>)

unités pertinentes pour le calcul du nombre de voyelles métriques (12)

type="vs"
voyelle stable

type="em"
"e" masculin

type="ef"
"e" féminin

type="ee"
"e" éli­dé

type="ei"
"e" igno­ré

type="sc"
semi-consonne

Quee Mars bêe et sanglant, quee Dianee bourre,

em vs vs vs vs vs em vs vs ef vs vs

1 2 3 4 5 6 7 8 9 10 11 12

Les différentes unités du calcul métrique (attribut de la balise <seg>)

unités pertinentes pour le calcul du mètre (6-6)

type="vs"
voyelle stable

type="em"
"e" masculin

type="ef"
"e" féminin

type="ee"
"e" élidé

type="ei"
"e" ignoré

type="sc"
semi-consonne

distribution des propriétés métricométriques

Unités prosodiques et morphosyntaxiques

1. [*Je viens dans son temple***e**] [*pour prier l'Eternel*]
2. [*Oui je viens dans sa***mai**] [*son pour voir Adeline*]
3. [*Viens dans son temple***pour**] [*adorer l'Eternel*]
4. [*Oui, je venais et***je**] [*vénérais l'Eternel*]
5. [*Oui, je viendrai dans***son**] [*temple pour l'adorer*]

Éléments de bibliographie

Beaudouin V. (2002) : *Mètre et rythmes du vers classique : Corneille et Racine*. Paris, Champion, coll. Lettres Numériques.

Cornulier de, B. (1982) : *Théorie du vers. Rimbaud, Verlaine, Mallarmé*, Paris, Seuil, coll. 'Travaux Linguistiques'.

Cornulier de, B. (1995) : *Art Poétique : Notions et problèmes de métrique*, Presses Universitaires de Lyon.