

Cette notice est la mise à jour de la présentation du programme qui se trouve dans :

G.M. ROCCATI, *A propos de la scansion automatisée des mètres latins: un programme spécialement conçu pour les textes médiévaux et humanistes*, in "Revue. Informatique et Statistique dans les Sciences Humaines (C.I.P.L.-L.A.S.L.A.)", t. XXVI (1990), pp. 221-231.

A. Nature et buts du programme – Le programme a un caractère essentiellement pragmatique. Il n'entend pas fournir un traitement complètement automatique, mais un instrument capable d'exécuter les tâches répétitives qui peuvent être formalisées dans le travail de scansion, ce qui permet de disposer de manière relativement rapide d'un « matériau » de travail qui peut être important : vers scandés, tables des formes métriques, index des termes avec indication des quantités prosodiques. Ce matériau de base est enregistré dans des fichiers lisibles par traitement de texte et dans des bases de données au format Microsoft Excel. En outre, pour réduire au minimum le travail manuel, le programme est conçu pour traiter presque directement un texte non préparé. Seules contraintes : en ce qui concerne la graphie, la normalisation *u/v* selon l'usage moderne est nécessaire ; l'enregistrement doit être en clair, en caractères ASCII, sans codes, c'est-à-dire en format « txt », en caractères non proportionnels (p. ex. *Courier*). Ne doivent pas non plus apparaître à l'intérieur du texte des renvois aux notes ou autres signes hétérogènes par rapport à la ligne à scander. Il faut donc, le cas échéant, éliminer préalablement les titres, lignes blanches, rubriques, indications de feuillets, etc. Le programme ne pourrait pas les reconnaître et échouerait évidemment dans sa tentative de traitement. En revanche sont admis dans le texte les signes de ponctuation, les tabulations, les blancs ou la numérotation des vers (à condition qu'ils se trouvent au début des vers). Il est donc possible, par exemple, d'utiliser directement, ou presque, le texte enregistré d'une édition.

Le programme se veut flexible, pour permettre de résoudre en cours de traitement les éventuels problèmes nouveaux qui pourraient se poser. C'est pourquoi le traitement est organisé en phases successives. Il est possible après chaque phase de parcourir à l'écran, et d'imprimer pour vérification, les résultats provisoires obtenus, de même qu'il est possible à chaque fois de recommencer le traitement à partir de l'une quelconque des phases déjà exécutées une fois introduites dans les textes traités les modifications ou les corrections qui se seraient révélées nécessaires. Les modifications peuvent être introduites soit dans le texte lui-même (pour corriger une erreur d'entrée, par exemple), soit dans la scansion proposée par la machine dans les cas où, pour une raison quelconque (texte dont le traitement est impossible ou erroné, licence poétique ou exception prosodique), la procédure ne peut fournir une scansion satisfaisante.

B. Phases du traitement – La scansion des textes se fait par une série de procédures automatiques ou semi-automatiques, en appliquant un nombre limité de règles, conçues comme des hypothèses de travail, vérifiées lorsqu'elles permettent la scansion. Les phases successives dans lesquelles le traitement s'articule sont regroupées sous les menus suivants : 1. *Général*; 2. *Opérations de base*; 3. *Scansion*; 4. *Classement des formes métriques*; 5. *Thesaurus lexical*; 6. *Procédures obsolètes*; 7. *Césure (hexamètre)*.

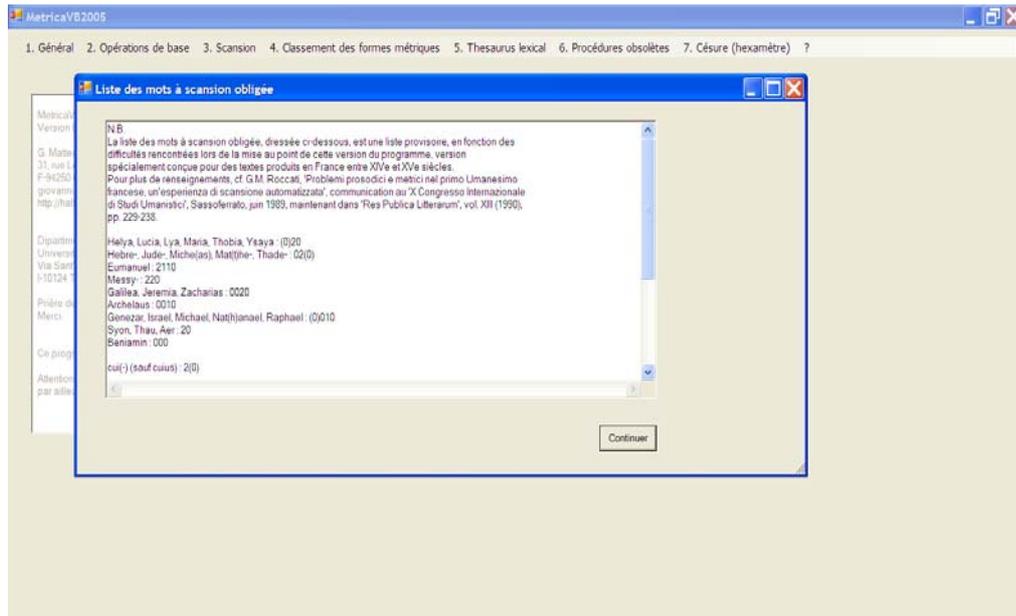


B.1 Première phase – La première phase (2.1 : le premier chiffre renvoie au menu, le deuxième à la procédure,) enregistre sous forme normalisée le texte à traiter et calcule les quantités de base, indépendamment du type de vers que l'on veut scander. Le calcul des quantités de base est exécuté par la création de séquences de caractères, invisibles pour l'utilisateur, reproduisant en forme normalisée, en majuscules sans ponctuation et en chiffres, le vers à scander. A chaque syllabe est attribuée une quantité: 0 = indéterminée, 1 = brève, 2 = longue. Les règles classiques sont appliquées, mais dans une forme simplifiée au maximum.

Les éléments suivants sont identifiés et la quantité attribuée:

- a. diphtongues (*oe, ae, au*);
- b. *u* et *i* semiconsonne (*qu, gu, i/j*);
- c. *h* intervocalique;
- d. brèves (voyelle suivie de voyelle) et longues (syllabe fermée) par position;
- e. monosyllabes se terminant en voyelle, toujours longs;
- f. *u* final, toujours long.

Un certain nombre de séquences particulières, dont la scansion serait aberrante si on appliquait mécaniquement ces règles, ont été enregistrées dans une table lors de la phase 2.1; la procédure consulte cette table et, avant tout autre calcul, introduit les quantités mémorisées.



Élision et hiatus sont également calculés. Sous la désignation d'« élision », il faut en fait entendre ici toute élimination de syllabe entre deux mots dans la scansion. Les règles latines ont été radicalement simplifiées: devant un début de mot vocalique, voyelle finale et syllabe finale se terminant en *m* sont toujours supprimées dans le décompte des quantités: cela revient à assimiler à l'élision des phénomènes tels que la synalèphe et l'aphérèse, bien différents dans la métrique classique. A la demande, le programme calcule aussi l'élision en dernière position pour le vers hypermètre. Si dans le vers il est possible d'élider un ou plusieurs mots, le programme crée autant de vers qu'il le faut pour réaliser les différentes possibilités: deux vers (2^1) pour une élision, quatre (2^2) pour deux, huit (2^3) pour trois, et ainsi de suite :

Josephina, 2^a distinctio (incipit: *Membra*), v. 129:
trois élisions possibles, donc huit vers
(résultat après phases 2.1 et 3.1.1)

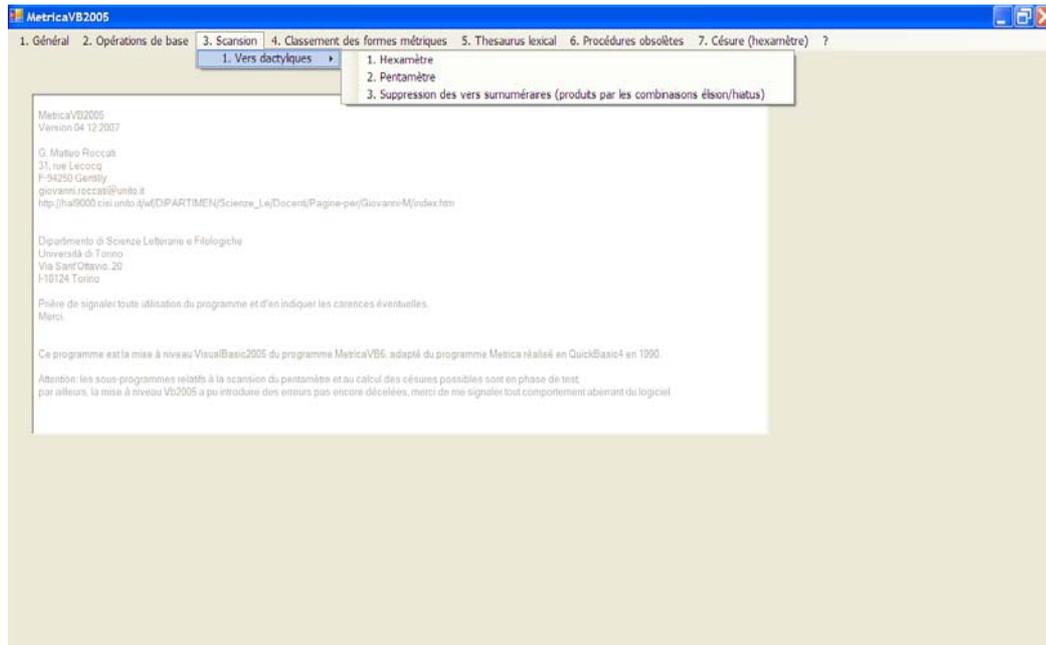
```

Membra / 129 / 1
Tristis utrumque genu atque ipsa incunabula nati
  2 1 1 2 1 1 2 2 2 2 1 1 2 0
Membra / 129 / 2
Tristis utrumque genu atque ipsa incunabula nati
  2 0 0 2 0 0 0 2 0 2 0 0 0 0 0 0
Membra / 129 / 3
Tristis utrumque genu atque ipsa incunabula nati
  2 0 0 2 0 0 2 2 2 2 2 2 1 1 2 0
Membra / 129 / 4
Tristis utrumque genu atque ipsa incunabula nati
  2 0 0 2 0 0 0 2 2 2 2 2 1 1 2 0
Membra / 129 / 5
Tristis utrumque genu atque ipsa incunabula nati
  2 0 0 2 2 2 2 2 2 2 2 2 1 1 2 0
Membra / 129 / 6
Tristis utrumque genu atque ipsa incunabula nati
  2 0 0 2 0 0 2 2 2 2 2 2 1 1 2 0
Membra / 129 / 7
Tristis utrumque genu atque ipsa incunabula nati
  2 0 0 2 2 2 2 2 2 2 2 2 1 1 2 0

```

Membra / 129 / 8
 Tristis utrumque genu atque ipsa incunabula nati
 2 0 0 2 0 0 2 2 2 2 2 2 1 1 2 0

B.2 Deuxième phase – La deuxième phase utilise la séquence des quantités ainsi déterminées pour scander complètement un type de vers particulier. Actuellement sont disponibles les procédures pour l'hexamètre (3.1.1) – avec possibilité de laisser libre choix ou d'imposer un dactyle en cinquième position – et pour le pentamètre (3.1.2).



Cette deuxième phase se décompose en plusieurs procédures, répétées jusqu'à ce qu'aucune intervention ne soit plus possible ; voici, à titre d'exemple, celles qui sont exécutées pour l'hexamètre :

- a. affectation de certaines valeurs obligatoires: 2 en première syllabe du vers et en première syllabe du sixième pied; 211 en cinquième position en cas de dactyle obligatoire;
- b. reconnaissance de séquences remarquables en position fixe: premier pied, cinquième pied en option dactylique prioritaire;
- c. reconnaissance de certaines séquences remarquables dans lesquelles une seule solution est possible: 210 > 211, 202 > 222, 012 > 112, etc.;
- d. reconnaissance des demi-pieds en position forte: calcul à partir du début et à partir de la fin du vers (chaque pied ayant un "poids" de 4 (22 ou 211), il est aisé de calculer les pieds et les demi-pieds entièrement scandés; un demi-pied fort a nécessairement valeur 2);
- e. répartition du "poids" restant sur la totalité du vers, lorsqu'une seule solution (1-2) est possible; la procédure calcule le "poids" total des quantités déterminées (a), le nombre de 0 (b), et répartit la valeur restante à déterminer: $x = (22 [5 \text{ pieds et demi}] - a) / b$;
- f. comparaison avec les différentes réalisations possibles en fonction du nombre de syllabes (les scansions possibles pour un hexamètre sont limitées; en admettant que dans les cinq premiers pieds on puisse avoir indifféremment un dactyle ou un spondée, elles sont égales à $32 (2^5)$). Si le cinquième pied est obligatoirement

un dactyle, elles se réduisent à 16 (2⁴): le vers est scandé intégralement seulement quand il existe une seule possibilité de scansion complète; si différentes scansions sont possibles ou si aucune ne l'est, le programme n'impose pas de scansion.

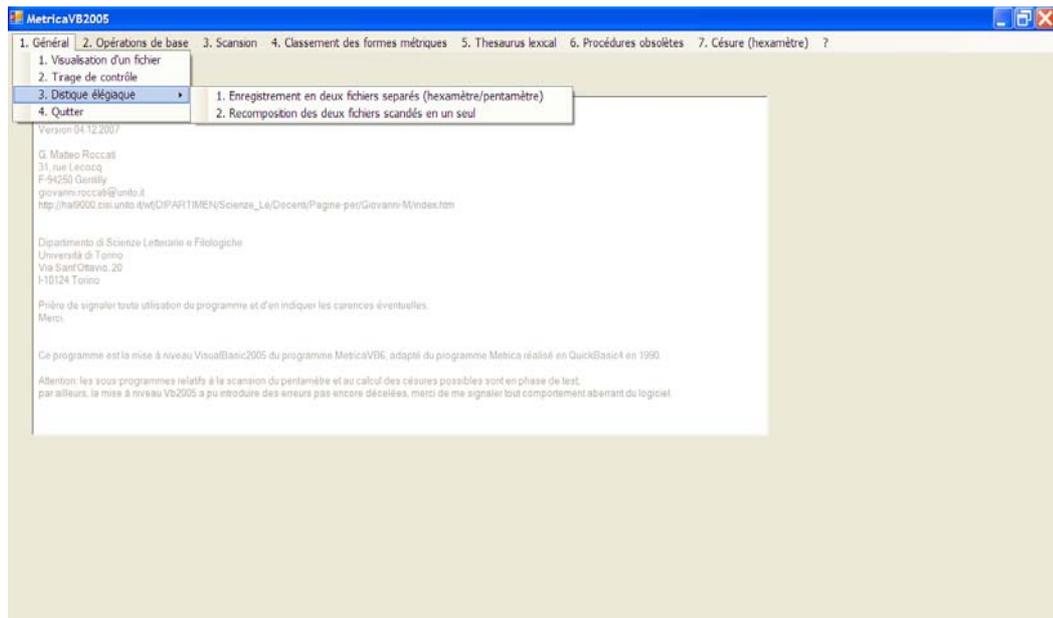
Lors de l'impression, ou à l'écran dans l'aperçu, pour faciliter la vérification de la scansion, cette phase fournit en annexe la liste des vers non conformes aux schémas métriques possibles et celle des vers qu'il n'a pas été possible de scander complètement.

B.3 Autres phases – Ces deux premières phases permettent de scander un pourcentage assez élevé de vers, de l'ordre de 75%, mais non la totalité. Pour rendre plus rapide l'introduction de quelques quantités sans être obligés de passer à travers un traitement de texte, une troisième phase (2.2) permet d'insérer manuellement les quantités que le programme n'a pas pu déterminer. Les vers contenant des quantités indéterminées (0) sont affichés à l'écran. En introduisant peu de quantités, évidentes à la lecture (diphthongues en *e*, *cas*, etc.) et en repassant une nouvelle fois par la deuxième phase, la presque totalité des vers peut être scandée.

Une quatrième phase (3.1.3) supprime les vers en surnombre créés par les différentes combinaisons possibles d'élision et hiatus afin que les résultats du traitement ultérieur, statistique ou lexical, ne soient pas faussés par la présence de ces vers fictifs. Le choix se fait en fonction des critères suivants: lorsqu'il a été possible de scander entièrement une seule des différentes possibilités, c'est cette dernière qui est retenue; dans les autres cas l'élision est privilégiée et le programme choisit la plus courte séquence de quantités.

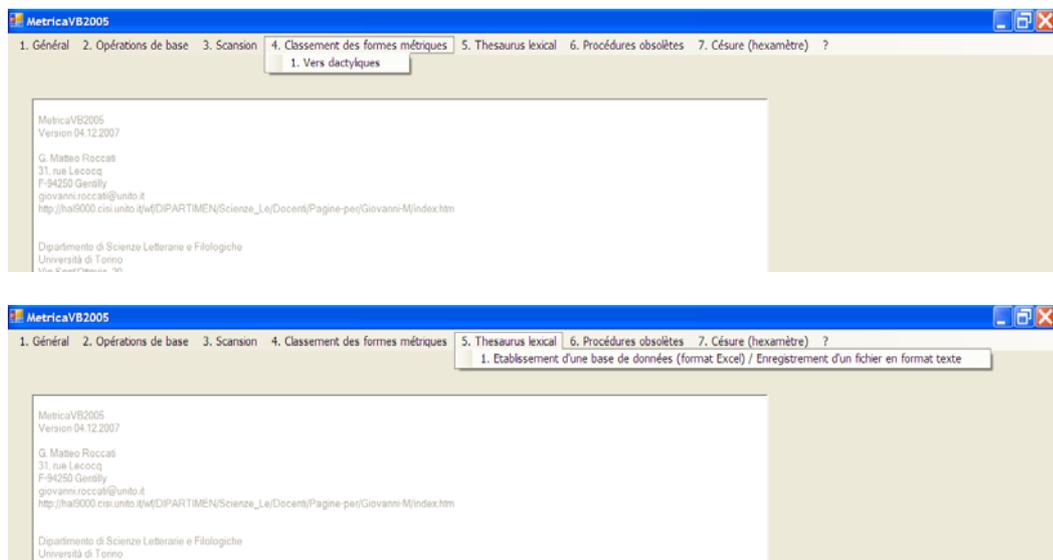
En interrompant la succession des différentes phases, il est possible à tout moment d'enregistrer les vers sous une forme lisible par n'importe quel traitement de texte (2.3); on introduit alors les corrections de quantité (synérèse, diérèse, etc.) que le programme ne peut pas déterminer et qui sont nécessaires pour une scansion correcte. On peut également à ce stade intégrer par exemple les variantes de la tradition manuscrite, éventuellement en gardant les différentes possibilités de scansion. Dans l'index des mots comportant la quantité prosodique, il suffira de modifier légèrement l'identification du vers (en ajoutant un numéro au premier mot de l'incipit: par exemple *Fontem* et *Fontem1*) pour pouvoir garder les différentes possibilités sans risque d'ambiguïté. Par contre il faudra évidemment tenir compte de tels rajouts et les éliminer si l'on veut réaliser des calculs statistiques. Ensuite, une deuxième procédure (2.4) permet d'enregistrer à nouveau un texte lisible par le programme.

Pour le pentamètre, une procédure permet de séparer le fichier texte de départ en deux fichiers, l'un comportant les hexamètres, l'autre les pentamètres (1.3.1), fichiers qui sont ensuite traités indépendamment l'un de l'autre (3.1.1 et 3.1.2). Une deuxième procédure (1.3.2) reconstitue, une fois la scansion achevée, le texte complet.



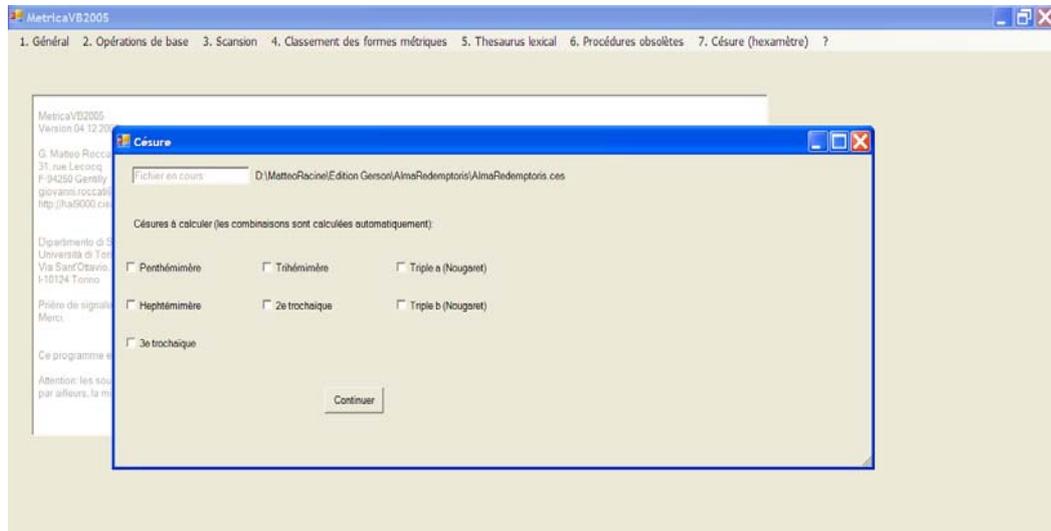
Par ailleurs, deux procédures permettent de visualiser à l'écran les fichiers de travail (1.1) et de les imprimer (1.2).

Lorsque le texte a été scandé complètement, il est possible d'utiliser les données mémorisées dans deux perspectives. La première (4.1) classe les différentes réalisations métriques en fonction de leur schème et des fréquences. Elle fournit la liste des scansion (avec renvoi aux vers où elles apparaissent) et les tableaux récapitulatifs (avec pourcentages). La deuxième (5.1) crée un index des formes lexicales, index qui comporte les quantités prosodiques et l'indication des vers dans lesquels elles apparaissent, en mettant ainsi en évidence constantes et fluctuations.



Enfin une dernière procédure (7), qui s'exécute également sur les vers scandés, permet de repérer les emplacement des césures. Le but de cette procédure – contrairement à celles qui visent l'établissement de la scansion – n'est pas de déterminer les césures « réelles », mais d'identifier les emplacements où la césure est possible ; naturellement plusieurs solutions peuvent être en concurrence dans le vers, seule la lecture et l'interprétation du texte permettent

de déterminer celles qu'il faut retenir. La procédure calcule l'emplacement des césures possibles suivantes (ainsi que certaines de leurs combinaisons) :



Pour faciliter la vérification, des tableaux récapitulatifs (lisibles par traitement de texte et en format Excel) indiquent les vers où les césures sont susceptibles d'apparaître. Les pourcentages se réfèrent au total des vers (vers avec césure x / total + vers sans césure x / total = 100%), la présence d'une césure n'étant pas exclusive de celle des autres, le total de la colonne sera toujours supérieur à 100%.

Dans le menu 6 (*Procédures obsolètes*) a été gardée la trace de procédures qui fonctionnaient dans les précédentes versions du programme et qui feraient maintenant double emploi avec celles des menus 4 et 5. Leur trace a été gardée car en revenant aux anciennes versions du programme elles peuvent éventuellement être utiles pour vérification.

Enfin un menu d'aide (?) indique rapidement les caractéristiques et les exigences de chaque procédure.

C. Limites du traitement – Un certain nombre d'"effets pervers" peuvent se réaliser lors de l'exécution. Afin de pouvoir intervenir, très souvent manuellement, et de corriger les fautes de scansion qui sont à imputer à la procédure et non au texte, il faut essayer de déterminer les cas où des "effets pervers" peuvent se produire, le plus fréquent se vérifiant lors de l'exécution de la routine "élision/hiatus". Il arrive en effet de manière relativement fréquente (5% des cas environ) que l'introduction d'un hiatus permette à la procédure "remplissage par division du 'poids' restant", d'affecter des valeurs à toutes les syllabes encore indéterminées, alors que, avec une élision, scansion en fait généralement correcte, la routine ne s'applique pas. Lors de l'élimination des vers surnuméraires, c'est la scansion complète, bien que fautive en réalité, qui sera retenue.

Exemple d'"effet pervers": la phase 3.1.1 scande complètement et de manière fautive le vers sans élision; la phase 3.1.3 le retiendra au détriment de la scansion correcte (222112221121120)
(*Josephina*, 2^a *distinctio*, v. 385).

Membra / 385 / 1
Maiori cumulo et ne gloria vexet inanis".
2 0 0 0 0 2 2 0 10 2 1 1 2 0

Membra / 385 / 2
 Maiori cumulo et ne gloria vexet inanis".
 2 1 1 2 1 1 2 2 2 11 2 1 1 2 0

Pour pallier ce genre de problèmes, il faut recourir à quelques vérifications, afin d'éliminer manuellement les erreurs de scansion. Le programme lui-même fournit un certain nombre d'instruments facilitant ces vérifications.

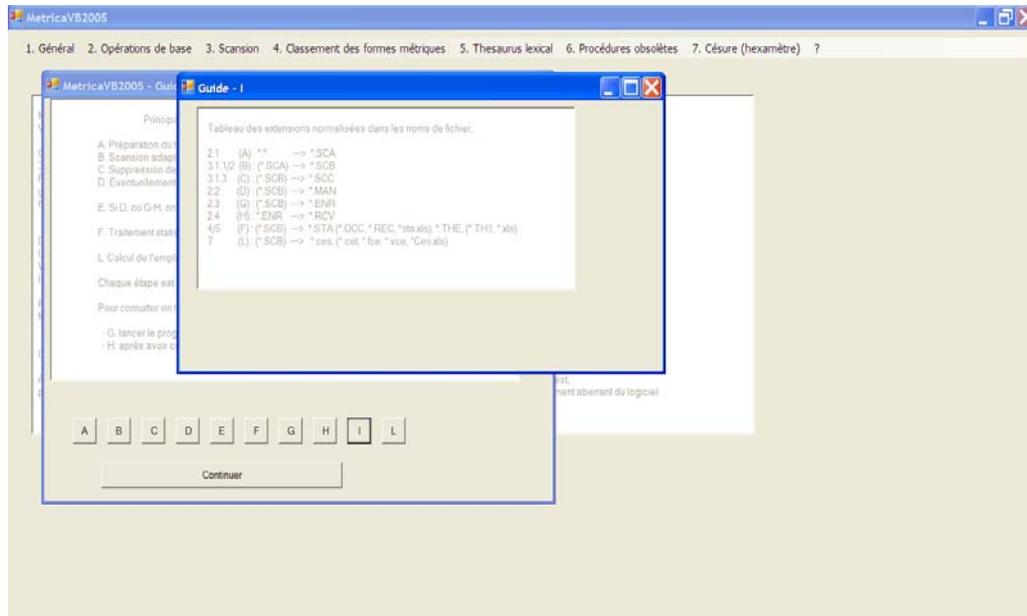
Lors de l'exécution de la routine "élision/hiatus", à chaque possibilité de scansion est affecté un numéro d'ordre qui s'ajoute au numéro du vers. Le n° 1 est toujours affecté à la possibilité ne comportant que des élisions. Il est donc facile, en parcourant la sortie imprimée (ou en faisant défiler le texte à l'écran), de repérer les vers où une scansion comportant au moins un hiatus a été préférée. Il suffit alors de vérifier les différentes possibilités de scansion produites par la phase 2.1 pour se rendre compte du bien fondé du choix de la machine. Le cas échéant il suffira de corriger directement la scansion erronée.

D'une manière générale, il est toujours possible que se produisent des scansions aberrantes, en particulier, par exemple, dans les vers où apparaissent des noms propres ou des termes grecs ou bibliques. L'utilisateur dispose de plusieurs moyens de vérification. Le premier est constitué par la liste fournie en annexe lors de l'exécution de la phase 3.1.1 (et 3.1.2). Comme on l'a vu, dans cette liste sont relevés tous les schémas métriques qui ne sont pas conformes au modèle donné et apparaissent les numéros des vers comportant des quantités indéterminées ou des séquences non prévues par la grille: par exemple les séquences 212 ou 111, ou encore des vers trop courts ou trop longs par rapport aux grilles possibles. Il faut évidemment vérifier alors manuellement les difficultés rencontrées par la machine.

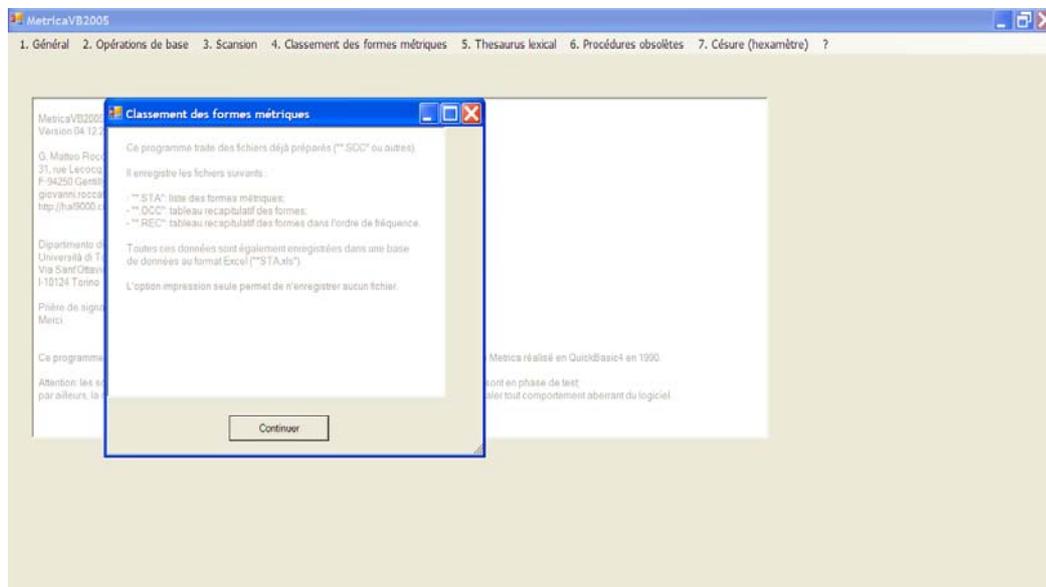
Un autre moyen de vérification des scansions est constitué par les tables des réalisations métriques fournies par la phase 4.1. Lorsque la masse de vers traités est suffisamment importante, il est très facile de repérer les schèmes de scansion uniques ou représentées par très peu de vers: il est nécessaire de vérifier ces cas, très souvent il s'agit de scansions aberrantes ou incomplètes.

Enfin l'index des termes fournit un moyen de vérification très utile. Cet index relève toutes les formes et établit un *item* chaque fois que la scansion diffère, même lorsqu'il s'agit du même terme. Les scansions aberrantes sont ainsi aisément repérables et il est très facile de se reporter alors pour vérification aux vers où elles apparaissent.

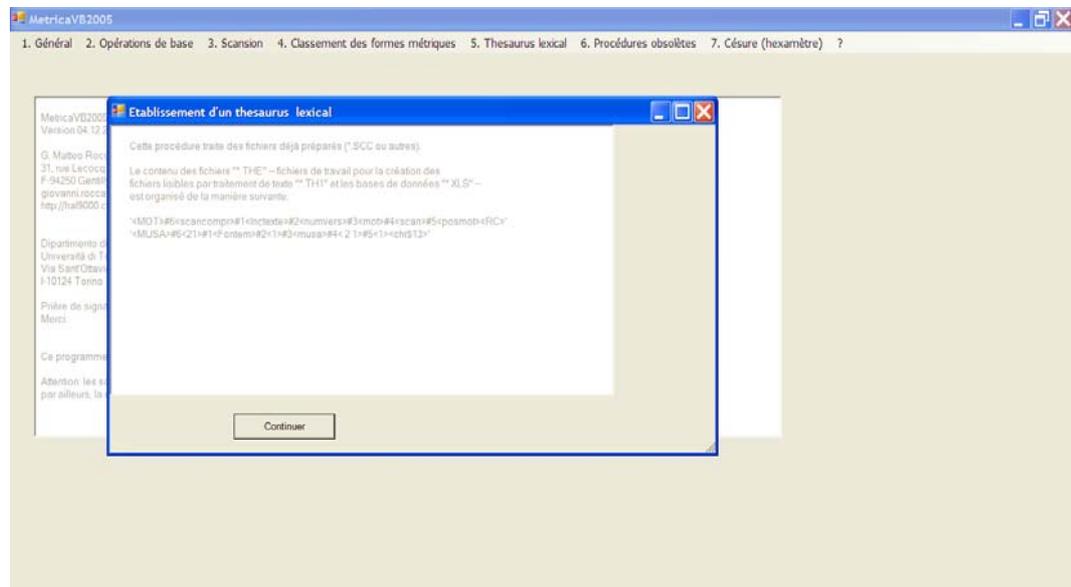
D. Extensions normalisées des fichiers.



Détail des fichiers enregistrés dans la procédure de classement des formes métriques:



Détail des fichiers enregistrés dans la procédure d'établissement de l'index lexical:



Détail des fichiers enregistrés dans le calcul de l'emplacement des césures:

- *.ces : le fichier lisible par le programme (et convertible en fichier lisible par traitement de texte) ;
- *.cet : le fichier *.ces, mais lisible par traitement de texte ;
- *.fce : fichier lisible par traitement de texte avec l'indication des vers où apparaissent les césures demandées ;
- *.vce : fichier lisible par traitement de texte avec le tableau récapitulatif des césures possibles ;
- *Ces.xls : base en format Excel comprenant le tableau récapitulatif, l'indication des vers où apparaît la césure demandée, la liste des vers avec l'indication des césures possibles.