

TACT

Un logiciel pour l'analyse littéraire et thématique

Le logiciel *TACT*¹ (*Thematic Analysis and Concordance of Texts*) a été développé à l'Université de Toronto par une équipe sous la direction de John Bradley en collaboration avec Lidio Presutti comme programmeur et en consultation avec Ian Lancashire du *Center for Computing in the Humanities*. L'Université de Toronto distribue le logiciel, tous droits réservés, sous condition que les utilisateurs ne modifient aucunement ses éléments constitutifs essentiels.

Ce logiciel offre de puissantes possibilités aux chercheurs qui s'intéressent à l'analyse quantitative de la littérature. Pour débiter, il faut avoir son texte sous forme électronique, le codage ASCII étendu étant le système de "défaut". L'utilisateur peut, s'il le veut, modifier l'alphabet (par exemple, ajouter un "å" pour le norvégien), les caractères (par exemple, lier le "ch" à la manière espagnole), et les marques diacritiques. Dans l'usage le plus élémentaire on passerait ce texte au logiciel *MAKBAS*, qui transforme le texte cru en une base de données textuelle, numérotant chaque ligne et créant un index où chaque mot est localisé selon le numéro de sa ligne. C'est en exploitant ce fichier que *TACT* peut faire une analyse interactive. Ce fichier, qui n'est pas un texte mais une base de données, prend beaucoup de place. Par exemple, d'une sélection de poèmes de Maurice Scève de 121K (*SCEVE.TXT*) *TACT* créa *SCEVE.TDB* de 402K! en à peu près dix minutes. Il s'agissait de 3.210 lignes contenant 20.741 mots (angl. : *tokens*) et 4.083 formes lexicales (angl. : *types*). Il faut garder le texte dans ces deux formes, parce qu'il est normal de vouloir revenir faire des rajustements dans les renseignements encodés du texte primaire. Le manuel de *TACT* propose même le cas où on voudrait deux bases de données textuelles différentes pour le même ouvrage; à titre

1. S'adresser à : *Center for Computing in the Humanities*, Roberts Library, 14th Floor, University of Toronto, Toronto, Ont, CANADA M5S 1A1. Téléphone : (416) 978-8656. On demande \$30 CAN ou \$25 US pour le manuel et le disque. Contexte informatique : micro-processeur type IBM-PC ou IBM-PS, système d'exploitation MS DOS 2.0 ou plus récent. Envoyer une disquette taille 3 1/2 in. *TACT* comporte plusieurs programmes, une documentation en anglais de plus de 500K, une sélection de *Volpone* à employer pour se familiariser avec le logiciel, tout sous forme comprimée (angl. : *archived*). *TACT* s'installe sur un disque dur.

d'exemple, un texte marqué pour des matières linguistiques et un autre pour les matières littéraires.

Pour revenir à notre cas le plus simple : *TACT* présente au chercheur une liste des mots employés et leur fréquence dans le texte. Pour mettre *TACT* en marche, l'utilisateur sélectionne le mot ou les mots qu'il veut étudier — par exemple les mots *absence* et *présence*.

absconsant.	1
absconse.	1
> absence	7
> absent	2
> absente	5
absince	1
absoudre.	1
abusai.	1
abuse	1
abuser.	1
abuserais	1
accélère.	1
acceptant	1
accident.	1

ECRAN N° 1: Section de la liste des mots,
dont trois sélectionnés²

2. Les écrans donnés ici sont réduits, soit en longueur, soit en largeur, mais autrement ils sont fidèles aux renseignements fournis par *TACT*.

TACT trouvera les emplacements de ces mots et les visualisera sous une des quatre formes maîtresses. Dans la forme "Text" le logiciel donne un écran de texte avec le mot en question mis en vedette (en l'occurrence en lettres jaunes).

```

Autant de fois pleine **elle nous est décrué,
Et deux Soleils, qui me'ont ci rencontré,
Autant de toi **ils me'ont la mémoire crue
Que me'est la force en la'attente recrué
Pour le long temps qui tant nous désassemble
Que vie et moi ne pouvons être ensemble.
> Car le mourir en cette longue absence
  - Non toutefois sans vivre en toi - me semble
  Service égal au souffrir en présence.

<d 36>
Le Forgeron vilainement erra,
Combien que'il sût telle être sa coutume

```

ECRAN N° 2: Visualisation "Text" (On note a que le
texte a subi plusieurs traitements avant d'être indexé.)

Dans la forme "KWIC" (*Key Word in Context*) on voit chaque usage des mots accompagné de deux lignes de texte qui précèdent et de deux qui suivent.

```
> absence (7)
Pour le long temps qui tant nous désassemble
Que vie et moi ne pouvons être ensemble.
Car le mourir en cette longue absence
- Non toutefois sans vivre en toi -- me semble
Service égal au souffrir en présence.
```

```
(35,8)
Que'aux bas mortels vient la froide terreur,
Qui de la peur de leur fin les offense.
Vois: seulement la mémoire en l'absence.
De toi me'échauffe et ard si vivement
Que'en toi me fait ta divine présence
```

```
(62,7)
Le jour passée de ta douce présence
Fut un serein en hiver ténébreux,
Qui fait prouver la nuit de ton absence
A le'œil de la'âme être un temps plus ombreux
Que n'est au Corps ce mien vivre encombreux
```

```
(129 3)
```

ECRAN N° 3: Visualisation en "KWIC"

Dans la forme "Index" on voit une liste qui fournit une seule ligne de texte pour chaque usage.

```
absence (7)
(35,8) | Car le mourir en cette longue >absence | - Non
        |                                         | toute
(62,7) | Vois: seulement la mémoire en la'>absence. | De
        |                                         | toi me'é
(129,3) | Qui fait prouver la nuit de ton >absence | A
        |                                         | le'œil de
(138,1) | Non tant me nuit cette si longue >absence, | Que
        |                                         | mal me
(141,9) | Parquoi de rien ne me nuit son >absence, | Vu
        |                                         | que'en t
(264,9) | Qui, malgré Mort, et malgré toute >absence, | Te
        |                                         | représen
(266,8) | bien que vous celait | Sa longue >absence, en
        |                                         | présence t

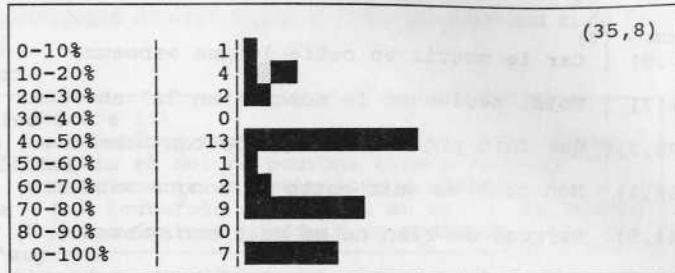
absent (2)
(144,4) | que **je soie, encore suis-je >absent. | Et si
        |                                         | nature
(215,4) | bonnement toutefois | Que, moi >absent, elle ne
        |                                         | soit

absente (5)
(46,9) | rets de servitude; | Plus je me'>absente, et plus
        |                                         | le ma
(144,1) | En toi je vis, où que tu sois >absente; | En moi
        |                                         | je m
(180,7) | Qui plusieurs fois du jugement se'>absente, |
        |                                         | Feignant du
(215,1) | | <d 215> | Je me'en >absente et tant et tan
(215,2) | Que'en la voyant, je la me cuide >absente; | Et
        |                                         | si **je

présence (14)
(6,4) | Se vit surpris de la douce >présence | Qui par sa
(35,10) | Service égal au souffrir en >présence. |
        |                                         | <d 36>
        |                                         | >
```

ECRAN N° 4: Visualisation en "Index" (On remarquera que sur sept usages d'"absence", six sont à la fin du vers!)

Enfin dans la forme "Distribution" on voit les occurrences présentées en graphique histogramme, qui laisse voir les endroits où le(s) mot(s) apparaissent dans le texte entier, par exemple si le(s) mot(s) sont concentrés au début ou à la fin du texte.



ECRAN N° 5: Visualisation "Distribution"

(Les chiffres en haut à droite identifient le bloc en relief.)

On a donc immédiatement à sa disposition une concordance, une liste de fréquences, un index du texte, et une présentation visuelle, comme géographique, de la distribution des listes sélectionnées.

Il faut remarquer que chacune de ces formes de présentation, sauf la forme "Text", peut être modifiée selon les besoins du chercheur. Voulez-vous voir dix lignes de texte pour chaque entrée dans le "KWIC", ou deux phrases avant et deux après l'usage en question? vous n'avez qu'à commander. Voulez-vous voir les usages d'*absence* et de *présence* arrangés selon leur séquence dans le texte? ou selon l'ordre alphabétique? choisissez.



ECRAN N° 6: Autre "Distribution" des mots de l'ECRAN N° 5

La puissance du logiciel réside dans sa nature interactive. L'utilisateur peut composer les instructions les plus complexes et *TACT* recherchera et extraira presque immédiatement les renseignements demandés. Ex. : on demande à voir tous les mots qui se terminent en *-eur*, ou tous les mots qui ont deux "c" (soit à côté, soit séparés), ou tous les mots qui ont seulement

la voyelle "a", ou même les demandes les plus fantaisistes : toutes les fois qu'un mot qui commence par "d" est suivi d'un mot de quatre lettres, puis d'un mot qui ne contient pas les lettres entre "m" et "t".

Criteria: AutoSelection Rule Menu
.*eur

```
> aigreur . . . . . 1
> architecteur. . . . . 1
> ardeur. . . . . 18
> blancheur . . . . . 3
> bonheur . . . . . 3
> chaleur . . . . . 2
> cœur . . . . . 61
> couleur . . . . . 5
> couronneur. . . . . 1
> donneur . . . . . 1
```

Criteria: AutoSelection Rule Menu
.*c.*c.*

```
> exercice. . . . . 1
> inaccessible. . . . . 1
> inconstance . . . . . 1
> inconvincible . . . . . 1
> innocence . . . . . 3
> licence . . . . . 2
> méconnaissance. . . . . 1
> obscurcit . . . . . 1
> occasion. . . . . 1
> occire. . . . . 1
```

Portions de deux ECRANS : règles pour les mots qui se terminent en *-eur* et ceux avec deux "c". Portions des listes des trouvés superposées.

Renseignements lexicaux : on demande la liste des mots à fréquence entre 10 et 25 dans le texte.

Criteria:		AutoSelection Rule Menu
freq 10:25		
> affection10	
> afin11	
> aimer11	
> air11	
> alors13	
> âme19	
> après10	
> arc16	
> archer12	
> ardent12	
Criteria:		
freq 25:45		
> ai35	
> ainsi38	
> ame40	
> aussi32	
> autre34	
> aux40	
> avec37	
> cette27	
> corps40	
> dame31	

Portions de deux ECRANS : les mots de fréquences 10 à 25 et 25 à 45.

La première est peut-être plus indicative sur le vocabulaire poétique de Scève.

Renseignements de position: on demande toutes les occurrences des mots *absence* et *présence* voisins (on peut varier la définition de "voisinage" — à cinquante mots de distance? à cent mots? sur la même page? dans la même phrase?). Renseignements thématiques : on peut associer plusieurs mots en un groupe (par exemple, toutes les formes d'un verbe, ou plusieurs mots qui désignent la mort) pour étudier leur disposition dans le texte. Toutes ces interrogations peuvent être formulées rapidement, et modifiées également rapidement. Les réponses aux enquêtes peuvent être sauveées en forme de fichier ou imprimées immédiatement, sous chacune des quatre formes mentionnées ci-dessus.

Mais la plus grande puissance de *TACT* reste à voir. Avant d'indexer le texte, l'utilisateur peut le parsemer d'étiquettes, de marques identificatrices anglais : *tags*), appelées *Ref ID* par *TACT*. Il est normal de situer un texte,

non par sa ligne, mais plutôt par son chapitre, sa page, son paragraphe peut-être, ou par actes et scènes dans une pièce. Le chercheur peut décider en quelles unités il veut diviser son texte. Alors il doit entrer ces indications (ex. : page 110) manuellement pour que *TACT* en tienne compte dans l'indexation. Un mot pourrait donc être localisé de plusieurs façons, selon les instructions du chercheur — page 110, paragraphe 2, ligne 10, ou plus simplement, page 110.

Mais il y a plus. Dans une pièce — ou dans un roman — on voudrait encoder d'autres renseignements. Par exemple, dans l'échantillon de *Volpone* fourni par *TACT* chaque discours est précédé par une marque qui donne le nom de l'interlocuteur. C'est l'utilisateur qui décide quel format donner à cette marque. Dans *Volpone* la notation *{*Nano*}* indique le début des paroles de Nano. A la fin de son discours une autre notation indiquera un changement de personnage ou une didascalie. (Remarquez un

danger: le mot NANO employé dans le texte pour désigner l'interlocuteur ne doit pas normalement être compté dans l'index, ni publié dans la concordance. *TACT* donne la possibilité d'une seconde notation pour indiquer ceci au logiciel. Un chercheur qui veut tenir compte d'une grande variété de phénomènes trouvera son texte alourdi par un assemblage de marques — utiles mais encombrantes.) Dans le *Volpone* on peut donc dresser une liste du vocabulaire de Nano, ou trouver ses emplois du mot "gold", ou contraster son vocabulaire dans l'acte deux avec celui de l'acte trois.

Dans un roman il est souvent important de distinguer entre les passages de *récit* et ceux de *dialogues*. (Et même un critique audacieux tenterait de désigner une troisième forme — le style indirect libre — qui pose des problèmes d'interprétation et de stylistique.) *TACT* distinguera ces deux ou trois niveaux de style — à condition que le chercheur étiquette le texte avant l'indexation. Alors, à lui d'imaginer les questions à poser à *TACT*.

Toutes ces indications sont déterminées par l'utilisateur, qui établit dans une session préalable à l'indexation les étiquettes que le logiciel doit compter. Le logiciel *MAKBAS* permet au chercheur de définir son alphabet, la signification des marques de ponctuation, les marques diacritiques, les divisions

du texte. On peut adapter un système de signes déjà connu, par exemple celui du *Oxford Concordance Program*. De tout ceci, *TACT* crée un logiciel *VOLPONE.MKS* qui contient les instructions particulières à ce texte.

Pour l'étude des topoï, *TACT* fournit la capacité de créer des "catégories", qui sont des faisceaux de mots formant un thème. On choisit ses propres groupes — par exemple, le groupe *mort-corps* serait composé de "cadavre", "corps", "dépouille", "restes", "charogne", tandis que le groupe *mort-décès* aurait "décès", "fin", "mort", "trépas". *TACT* chercherait tous les usages du groupe entier, selon les instructions du chercheur. On peut, naturellement, éliminer les usages qui ne font pas partie du groupe ("fin" qui veut dire "subtil", *et cetera*). On peut utiliser cette capacité du logiciel pour lemmatiser la liste des mots du texte. Une fois qu'on a créé sa catégorie, elle peut être employée dans l'étude d'autres textes — et modifiée selon chacun si nécessaire.

TACT est une œuvre en progrès. Ses créateurs ont l'intention d'ajouter d'autres capacités à un programme déjà très riche et extraordinairement flexible.

Craig B. Brush
Université Fordham