

## ANÁLISE MECÂNICA DO TEXTO (TACT)

*António Rodrigues de Almeida\**

A máquina que é o computador é hoje largamente reconhecida como um instrumento útil para a área das Letras. A sua utilização, em já vastas regiões do globo, está levando as nossas disciplinas por um caminho que não aparece ainda nítido no seu traçado, mas que levará muito provavelmente e muito rapidamente a um novo mundo das letras.

Pretende-se com esta comunicação dar um contributo para a maior divulgação do TACT, um programa para PC que pensamos ser útil na análise do texto, tanto numa perspectiva literária como linguística. É um instrumento que foi criado para a nossa área, num meio académico como o nosso, embora mais a norte, que é acessível ao "pobre" das letras, e que é hoje usado e reconhecido como ferramenta útil num grande número de países mais avançados neste tipo de aplicações.

Não iremos falar propriamente da ANÁLISE MECÂNICA DO TEXTO, mas do TACT, que permite fazer a análise mecânica do texto. Partimos do princípio de que a maioria dos ouvintes não conhece o TACT e é, por isso, para eles que imaginámos esta oportunidade como um espaço de informação.

---

\* Professor da Faculdade de Letras da Universidade de Lisboa.

## O nome

TACT: Textual Analysis Computing Tools

## História

O TACT foi lançado pela primeira vez em Junho de 1989, em Toronto, por ocasião do Congresso da Association for Computers and the Humanities e da Association for Literary and Linguistic Computing. A versão 1.2 (A e B) saiu em 1991. A versão 2.0 foi preparada para o projecto *ADMYTE: Archivo Digital de Manuscritos y Textos Españoles* de Francisco Marcos-Marín, tendo em conta particularmente as características próprias dos textos em castelhano medieval.

A versão 2.1 representa uma remodelação substancial do programa, mormente no que se refere ao uso da memória e ao alargamento das suas possibilidades, e foi dada a conhecer em versão beta em Junho de 1993, estando presentemente na sua versão gamma. O Manual para a versão 2.1 será publicado pela MLA Modern Language Association of America e espera-se para os finais de 1994.

O programa é da responsabilidade do Centre for Computing in the Humanities, da Universidade de Toronto, que até 1989 contou com o apoio da IBM Canada. É director do CCH o Prof. Ian Lancashire.

## O TACT e outros programas

O TACT é formado por um conjunto de programas, alguns deles em parte inspirados por outros programas produzidos quer igualmente pela Universidade de Toronto (MTAS) quer por outras instituições (ARRAS e WordCruncher).

## Tarefas que podem ser executadas com o TACT.

1. investigar o vocabulário do texto (o texto usado para os exemplos é a *Eneida* de Virgílio), como:

(1) palavras em *-plex* (formas e respectivo número)

<i>duplex</i> (1)	<i>simplicis</i> (1)
<i>duplicem</i> (2)	<i>supplex</i> (18)
<i>duplices</i> (1)	<i>supplicibus</i> (2)
<i>duplicis</i> (4)	<i>supplicis</i> (1)
<i>multiplicem</i> (1)	<i>triplex</i> (1)
<i>septuplicis</i> (1)	

(2) palavras em *-tia* (formas e respectivo número)

<i>amicitiam</i> (1)	<i>laetitiae</i> (1)
<i>dementia</i> (2)	<i>laetitiam</i> (1)
<i>diuitiis</i> (1)	<i>militia</i> (2)
<i>gratia</i> (5)	<i>militiam</i> (1)
<i>hostia</i> (3)	<i>opulentia</i> (1)
<i>inclementia</i> (1)	<i>potentia</i> (3)
<i>inertia</i> (5)	<i>praesentia</i> (2)
<i>iustitia</i> (1)	<i>primitiae</i> (2)
<i>iustitiae</i> (2)	<i>prudentia</i> (1)
<i>iustitiam</i> (1)	<i>uiolentia</i> (4)
<i>laetitia</i> (6)	

2. criar listas de palavras com a indicação da sua frequência e do lugar onde ocorrem no texto, como:

(3) palavras em *-cola*

<i>accola</i> (1)	(E,7,729)
<i>agricola</i> (1)	(E,10,805)
<i>agricolae</i> (2)	(E,2,627; E,7,521)
<i>appenninicolae</i> (1)	(E,11,700)
<i>caelicolae</i> (5)	(E,2,640; E,6,554; E,10,6; E,10,97; E,10,117)
<i>caelicolas</i> (1)	(E,6,787)
<i>siluicolae</i> (1)	(E,10,551)

3. criar concordâncias, com a inserção das palavras nos seus contextos (a extensão destes pode ser maior ou menor, dependendo da escolha do utilizador):

(4) *agricolae* (2)

ac ueluti summis antiquam in montibus ornum  
 cum ferro accisam crebrisque bipennibus instant  
 eruere **agricolae** certatim, illa usque minatur  
 et tremefacta comam concusso uertice nutat,  
 uulneribus donec paulatim euicta supremum

(E,2,627)

tum uero ad uocem celeres, qua bucina signum  
 dira dedit, raptis concurrunt undique telis  
 indomiti **agricolae**, nec non et Troia pubes  
 Ascanio auxilium castris effundit apertis.  
 derexere acies. non iam certamine agresti

(E,7,521)

4. seleccionar grupos de palavras do texto, na dependência de uma grande variedade de condições; por exemplo, da:
- a) coocorrência de formas; exemplo:

(5) *ad* + palavras de 5 letras terminadas em *-am*:

(E,1,618)	Venus Phrygii genuit Simoentis	>ad undam?	atque equidem
(E,3,302)	urbem in luco falsi Simoentis	>ad undam	libabat cineri
(E,3,509)	optatae gremio telluris	>ad undam	sortiti remos
(E,4,513)	Auerni, l falcibus et messae	>ad lunam	quaeruntur aenis
(E,9,22)	in arma uocas.'et sic effatus	>ad undam	processit
	summoque		
(E,10,282)	laudes. ultro occurramus	>ad undam	dum trepidi
(E,11,327)	complere ualent, iacet omnis	>ad undam	materies: ipsi

- b) localização no texto (com maior ou menor afastamento relativo):

(6) *ad* + palavras de 5 letras terminadas em *-am* separadas por 1 palavra:

(E,3,389)	cum tibi sollicito secreti	>ad fluminis undam	litoreis
(E,6,714)	corpora debentur, Lethaei	>ad fluminis undam	securos
(E,10,833)	Interca genitor Tiberini	>ad fluminis undam	uulnera

- c) forma das palavras; assim, por exemplo, palavras começadas ou terminadas por uma dada letra, por uma ou mais letras, com uma dada extensão; exemplos:

## (7) formas de 15 letras:

*amphitryoniadae* (1) (E,8,103)  
*amphitryoniades* (1) (E,8,214)  
*consanguinitate* (1) (E,2,85)

(8) formas começadas por f e terminadas em *-tur*:

*fatentur* (2) (E,11,344; E,12,568)  
*feruntur* (7) (E,5,157; E,5,587; E,7,156; E,7,673; E,9,120;  
E,11,906; E,12,575)  
*fīrmantur* (1) (E,9,785)  
*franguntur* (1) (E,1,104)  
*fruuntur* (2) (E,3,352; E,8,468)  
*funduntur* (1) (E,6,709)

## d) função, significado ou estilo das palavras; exemplos:

(9) preposições (*a, ab, ad, de, per, pro*)

*a* (65)  
*ab* (134)  
*ad* (322)  
*de* (112)  
*per* (350)  
*pro* (49)

## (10) palavras que significam "dizer":

*ait* (45) *aiunt* (1) *dic* (6) *dicabo* (2) *dicam* (7) *dicamus* (1) *dicatis* (1)  
*dicens* (4) *dicent* (1) *dicente* (1) *dicentem* (1) *dicentur* (1) *dicere* (11)  
*diceret* (1) *dicet* (2) *dicetur* (1) *dici* (1) *dicit* (1) *dicite* (5) *dicitur* (6)  
*dicta* (42) *dictam* (1) *dicti* (1) *dictis* (61) *dicto* (5) *dictu* (11) *dictum* (1)  
*dicturus* (1) *dictus* (2) *dixera* (25) *dixere* (1) *dixerunt* (1) *diximus* (1)  
*dixisse* (2) *dixit* (31) *fare* (3) *fari* (6) *farier* (1) *inquit* (29) *locuta* (5)  
*locutam* (1) *locuti* (1) *locutus* (8) *loquar* (1) *loquendi* (1) *loquentis* (3)  
*loqui* (3) *loquitur* (1) *loquor* (2) *loquuntur* (1)

## (11) nomes de deuses:

*Aeole* (2) *Aeolus* (4) *Apolline* (1) *Apollinis* (3) *Apollo* (26) *Cupidine* (4)  
*Cupido* (11) *Cupidus* (1) *Diana* (2) *Dianae* (8) *Dianam* (1) *Hercule* (1)  
*Herculis* (3) *Iani* (1) *Ianum* (1) *Ianus* (2) *Ioue* (6) *louem* (9) *Ioui* (4)  
*Iouis* (32) *Iri* (1) *Irim* (5) *Iris* (2) *Iuno* (35) *Iunone* (1) *Iunonem* (3)  
*Iunoni* (6) *Iunonis* (11) *Iuppiter* (35) *Mars* (3) *Marte* (12) *Martem* (8)  
*Marti* (2) *Martis* (10) *Mercurio* (1) *Mercurium* (1) *Mercurius* (1)  
*Minerua* (1) *Mineruae* (8) *Mineruum* (1) *Venerem* (3) *Veneri* (3)  
*Veneris* (9) *Venus* (19)

(12) nomes geográficos referentes à península itálica:

*Ausonia* (6) *Ausoniae* (6) *Ausoniam* (2) *Ausonias* (3)  
*Hesperia* (1) *Hesperiae* (2) *Hesperiam* (10) *Italia* (4)  
*Italiae* (8) *Italiam* (32) *Oenotria* (1)

(13) palavras que significam cores:

*alba* (6) *albae* (1) *albam* (6) *albi* (2) *albis* (7) *albo*  
 (2) *albos* (3) *album* (1) *albus* (1) *nigra* (6) *nigrae* (1)  
*nigram* (1) *nigras* (1) *nigri* (2) *nigro* (3) *nigrum* (1)

(14) palavras poéticas (exemplo: nomes em *-ides*):

*Achaemenides* (2) *Aeacidae* (3) *Aeaciden* (1) *Aeneadae* (10) *Aeneadas*  
 (3) *Aeneadis* (1) *Aenide* (1) *Aeoliden* (2) *Aeolides* (1) *Alcidae* (2) *Alcide*  
 (1) *Alciden* (3) *Alcides* (6) *Aloidas* (1) *Amphitryoniadae* (1)  
*Amphitryoniades* (1) *Anchisiades* (4) *Antenoridas* (1) *Arcadas* (4)  
*Arcades* (9) *Atlantide* (1) *Atridae* (3) *Atridas* (3) *Atrides* (1) *Atridis* (1)  
*Aulide* (1) *Belidae* (1) *Cecropidae* (1) *Cycladas* (2) *Dardanidae* (15)  
*Dardanides* (2) *Dardanidis* (2) *Epytiden* (1) *Epytides* (1) *Eumenidum*  
 (4) *Eurytides* (1) *Garamantide* (1) *Haemonides* (1) *Hesperidum* (2)  
*Hippotaden* (1) *Hyadas* (2) *Hyrtacidae* (2) *Hyrtacides* (3) *Iaside* (1)  
*lasides* (1) *Iliades* (3) *Imbrasidas* (1) *Imbrasides* (1) *Italides* (1)  
*Lampadis* (1) *Laomedontiadae* (1) *Laomedontiaden* (2) *Libystidis* (2)  
*Maeonidae* (1) *Mygdonides* (1) *Occidis* (1) *Oreades* (1) *Othryadae* (1)  
*Othryades* (1) *Palladis* (7) *Pelidae* (3) *Pelides* (1) *Priamiden* (2)  
*Priamides* (2) *Romulidis* (1) *Scipiadas* (1) *Strophades* (1) *Thybridis* (3)  
*Tritonidis* (1) *Troades* (1) *Tydides* (6) *Tyndaridis* (1) *Tyrrhidae* (2)

(15) arcaísmos (exemplo: infinitivos em *-ier*):

*accingier* (1) *admittier* (1) *defendier* (1) *dominarier* (1) *farier* (1)

5. seleccionar palavras que convergem ou diferem pelo seu significado ou pela sua forma: por exemplo, palavras com dados prefixos (exemplos: palavras de 7 letras com prefixo *per-* e palavras com sufixo *-ger*):

(16) *peragat* (1) *peredit* (1) *peregat* (1) *pereunt* (1) *perfert* (2) *pergant* (1)  
*pergunt* (1) *peribat* (1) *perstat* (1)

(17) *aligeri* (1) *aligerum* (1) *armiger* (5) *armigerum* (1) *corniger* (1)  
*lanigerae* (1) *lanigeras* (2) *lanigeros* (1) *saetigeri* (2) *saetigeros* (1)  
*turrigerae* (2)

6. seleccionar as palavras em função da estrutura do texto (ex.: obra, livro, verso):

(18) coocorrência de *pius* e *Aeneas* e a sua distribuição na obra (note-se a sua posição fixa, após a primeira palavra do verso):

(E,1,220)	exaudire uocatos.   praecipue	>pius Aeneas nunc acris
(E,1,305)	mentemque benignam.   At	>pius Aeneas per noctem plurima
(E,1,378)	tempestas appulit oris.   sum	>pius Aeneas, raptos qui ex
(E,4,393)	stratisque reponunt.   At	>pius Aeneas, quamquam lenire
(E,5,26)	seruata remetiore astra.   tum	>pius Aeneas: 'equidem sic
(E,5,286)	sub ubere nati.   Hoc	>pius Aeneas misso certamine
(E,5,685)	flumina prosunt.   tum	>pius Aeneas umeris abscindere
(E,6,9)	flumina monstrat.   at	>pius Aeneas arces quibus altus
(E,6,176)	clamore fremebant,   praecipue	>pius Aeneas. tum iussa
(E,6,232)	dixitque nouissima uerba.   at	>pius Aeneas ingenti mole
(E,8,84)	litore conspicitur sus;   quam	>pius Aeneas tibi enim, tibi,
(E,9,253)	tum cetera reddet   actutum	>pius Aeneas atque integer
(E,10,591)	uoluit aruis.   quem	>pius Aeneas dictis adfatur
(E,10,783)	reminiscitur Argos.   tum	>pius Aeneas hastam iacit; illa
(E,10,826)	pro laudibus istis,   quid	>pius Aeneas tanta dabit indole
(E,11,170)	te funere, Palla,   quam	>pius Aeneas et quam magni
(E,12,175)	altaria libant.   Tum	>pius Aeneas stricto sic ense
(E,12,311)	conduntur lumina noctem.   At	>pius Aeneas dextram tendebat

7. distribuir as palavras pelo texto, permitindo ver se uma determinada forma tem uma distribuição homogénea ou não ao longo do texto:

(19) distribuição das formas em *-ntur*:

1		18 +++++
2		13 +++++
3		12 +++++
4		8 +++++
5		14 +++++
6		16 +++++
7		16 +++++
8		11 +++++
9		14 +++++
10		8 +++++
11		18 +++++
12		18 +++++

Total: 166

(20) distribuição de "Aeneas" ao longo da obra:

1		28 +++++
2		1 +
3		5 +++++
4		14 +++++
5		27 +++++
6		28 +++++
7		9 +++++
8		21 +++++
9		13 +++++
10		40 +++++
11		18 +++++
12		33 +++++

Total: 237

(21) distribuição das palavras que designam "deusas" ao longo da obra:

1		25 +++++
2		10 +++++
3		7 +++++
4		14 +++++
5		7 +++++
6		8 +++++
7		15 +++++
8		11 +++++
9		6 +++++
10		13 +++++
11		8 +++++
12		8 +++++

Total: 132

(22) distribuição das palavras que designam "deuses" ao longo da obra:

1		37 +++++
2		13 +++++
3		22 +++++
4		31 +++++
5		15 +++++
6		17 +++++
7		34 +++++
8		26 +++++
9		29 +++++
10		30 +++++
11		14 +++++
12		35 +++++

Total: 303



## Resultados

Os resultados destas análises são apresentadas sob diferentes visualizações (exemplo: palavras em *-bundus*):

1. em listas, com indicação da frequência:

(23) *furibunda* (2)  
*moribunda* (2)  
*moribundam* (1)  
*moribundum* (1)  
*moribundus* (1)  
*tremibunda* (1)

2. em índices, com indicação do lugar onde ocorrem:

(24) *furibunda* (2) (E,4,645; E,7,348)  
*moribunda* (2) (E,6,732; E,10,341)  
*moribundam* (1) (E,4,323)  
*moribundum* (1) (E,5,374)  
*moribundus* (1) (E,10,590)  
*tremibunda* (1) (E,10,522)

3. em contexto, com o texto antes e depois:

(25) contexto reduzido de tipo "concordância":

(E,4,323)	adibam,   fama prior. cui me	>moribundam deseris hospes
(E,4,645)	limina et altos   conscendit	>furibunda rogos ensemque
(E,5,374)	ferebat,   perculit et fulua	>moribundum extendit harena.
(E,6,732)	terrenique hebetant artus	>moribundaque membra.   hinc
(E,7,348)	ad intima subdit,   quo	>furibunda domum monstro
(E,10,341)	dexteraque ex umero neruis	>moribunda pependit.   tum
(E,10,522)	hastam:   ille astu subit, at	>tremibunda superuolat hasta,
(E,10,590)	inguen;   excussus curru	>moribundus uoluitur aruis.

(26) contexto maior ou menor (aqui: 3 versos):

*furibunda* (2)  
interiora domus inrumpit limina et altos  
conscendit **furibunda** rogos ensemque recludit  
Dardanium, non hos quaesitum munus in usus.

(E,4,645)

conicit, inque sinum praecordia ad intima subdit,  
 quo **furibunda** domum monstro permisceat omnem.  
 ille inter uestis et leuia pectora lapsus

(E,7,348)

moribunda (2)  
 seminibus, quantum non noxia corpora tardant  
 terrenique hebetant artus **moribunda**que membra.  
 hinc metuunt cupiuntque, dolent gaudentque, neque auras

(E,6,732)

protinus hasta fugit seruatque cruenta tenorem,  
 dexteraque ex umero neruis **moribunda** pependit.  
 tum Numitor iaculo fratris de corpore raptio

(E,10,341)

moribundam (1).  
 exstinctus pudor et, qua sola sidera adibam,  
 fama prior. cui me **moribundam** descriis hospes  
 (hoc solum nomen quoniam de coniuge restat)?

(E,4,323)

moribundum (1)  
 Bebrycia ueniens Amyci de gente ferebat,  
 percudit et fulua **moribundum** extendit harena.  
 talis prima Dares caput altum in proelia tollit,

(E,5,374)

moribundus (1)  
 fulgentis clipei, tum laeuum perforat inguen;  
 excussus curru **moribundus** uoluitur aruis.  
 quem pius Aeneas dictis adfatur amaris:

(E,10,590)

tremibunda (1)  
 inde Mago procul infensam contenderat hastam:  
 ille astu subit, at **tremibunda** superuolat hasta,  
 et genua amplectens effatur talia supplex:

(E,10,522)

4. em gráfico, com a sua distribuição no texto:

(26) por partes estruturais da obra (aqui: livros)

4		2l++
5		1l+
6		1l+
7		1l+
10		3l+++
Total: 8		

(27) por partes de 10% da extensão total da obra:

0-10%		0l
10-20%		0l
20-30%		2l++
30-40%		1l+
40-50%		1l+
50-60%		1l+
60-70%		0l
70-80%		3l+++
80-90%		0l
90-100%		0l

5. no texto, com a forma em evidência:

(28) uictorem Buten immani corpore, qui se  
 Bebrycia ueniens Amyci de gente ferebat,  
 perculit et fulua **moribundum** extendit harena.  
 talis prima Dares caput altum in proelia tollit,  
 ostenditque umeros latos alternaque iactat

Estas diferentes visualizações podem ser chamadas ao ecrã simultaneamente, podendo o utilizador ver ao mesmo tempo, por exemplo, a lista de formas e o gráfico da sua distribuição no texto.

Por outro lado, as diferentes visualizações estão ligadas, sendo possível, por isso, ir directamente de uma forma do texto para a posição que lhe corresponde no gráfico ou na concordância.

**O que é necessário para que o TACT execute as suas tarefas de análise.**

O TACT pressupõe um texto em ASCII, a partir do qual constrói uma BASE DE TEXTO, desta se servindo para fazer as suas análises. Para isso, usa diferentes subprogramas. Uma vez estabelecida a BASE DE TEXTO, o TACT não precisa do texto original para as suas análises.

**O texto.**

O texto é em ASCII e deve ser preparado pelo utilizador, em função das tarefas que pretende sejam realizadas pelo TACT, principalmente no que se refere a:

- a) formato ASCII;
- b) introdução das marcas relativas à estrutura do texto;
- c) introdução das marcas relativas às características de função, significado, estilo, etc.
  
- d) em latim:
  - separação das enclíticas
  - eventual distinção de i vogal / i consoante e de u vogal / u consoante
  - eventual marcação de particularidades linguísticas (ex.: quantidade, hiato, ...)

### **Extensão do texto.**

A BASE DE TEXTO pode ser feita a partir de um ou mais textos, pequenos ou grandes, que no seu conjunto podem atingir uma dimensão bastante grande.

### **Capacidade multilingue**

O TACT pode tratar textos em grande número de línguas, resolvendo com facilidade os problemas decorrentes do uso de diacríticos; com a ajuda de programas de fontes adequados, pode igualmente tratar textos que usam alfabetos não latinos, como o grego, embora não o árabe e o hebraico ou textos em caracteres ideográficos.

### **Os programas (versão 2.1)**

1. **MakeBase**, usado na construção da BASE DE TEXTO, no que pode ser complementado pelos dois programas seguintes.
2. **MergeBase**, que reúne o resultado de várias operações do **MakeBase**.
3. **BuildBat**, que automatiza as operações dos dois anteriores.
4. **UseBase**, usado na análise da BASE DE TEXTO.

5. **Collgen**, que faz o levantamento dos grupos de palavras repetidos.
6. **TACTStat**, que faz a estatística relativa à extensão e à frequência das formas.
7. **TACTFreq**, que produz listas das formas, por ordem alfabética, ordem inversa e ordem de frequências
8. **TACTDct**, que ajuda à criação de dicionários das formas de um texto e à marcação destas (lema, parte do discurso, categoria semântica ou outra, etc.)
9. **Anagrams**, que faz o levantamento das formas do texto que usam as mesmas letras. Ex.:

(29) *amor, Roma, mora, Maro, oram, armo*

## Bibliografia

1. JOHN BRADLEY, *TACT. User's guide*. Version 1.2. Toronto, University of Toronto, 1990
2. T. R. WOOLDRIDGE (ed.), *A Tact exemplar*. Toronto, University of Toronto, 1991
3. I. LANCASHIRE, *Humanities Computing: The CCH Toronto-IBM Canada Co-operative*. Toronto, University of Toronto, 1990
4. P. VERGILI MARONIS *Opera*, ed. R. A. B. MYNORS, O.U.P., 1969