

21.1.4 CONVENZIONI ED AVVERTENZE SPECIALI (CT VER. 1.8). Restano alcune convenzioni ed avvertenze che è bene avere ancora sott'occhio in una *quick reference guide* del CT come la presente.

In primo luogo sono da ricordare alcuni simboli usati sistematicamente nel CT, qui elencati insieme al loro "nome" tipografico internazionale, alla spiegazione del fenomeno che simboleggiano e ad un esempio dal corpus:

simboli	"nome"	fenomeno	esempio (e glosse)
¬	<i>logicalnot</i>	composti fonosintassi di n\d	Mei-di-donna nonn¬ è, foglia-d è
	<i>brokenbar</i>	corsivi filologici	or o
·	<i>periodcenter</i>	proclisi+assimilazione	de· regno, Be· ll' ho
÷	<i>divide</i>	grafoclisi	porta ÷l ÷te ÷ne
~	<i>tilde</i>	compendi	ka~ agosto
^	<i>caret</i>	ellissi	^^^ (corrisponde all'usuale "...")
[^]	<i>caret tra quadre</i>	lacuna	molto [^^^] francamente
×	<i>multiply</i>	deperdita	xxosta (caratteri illeggibili)
*	<i>asterisk</i>	vacua	die *** d' aprile (spazio bianco nel ms.)
Ø	<i>Oslash</i>	morfemi zero	a ÷Ø demoní ('ai demonii')
©...®	<i>copyright... register</i>	corsivi tipografici	le credenze del © Credo in Deo ®
(...)	<i>round brackets</i>	abbreviazioni sciolte	mante(n) gna
[...]	<i>square brackets</i>	integrazioni	rag[g]io
{...}	<i>brace brackets</i>	simboli grafici	{SN} ('signum notarii')

Tav. 254: Simboli convenzionali usati nel CT.

In secondo luogo è bene ricordare le seguenti quattro convenzioni:

- (j) i **grafoclitici** sono trattati come token individuali marcati dal *divide* (ASCII 246 = ANSI 247) "÷";
- (ij) i **lemmi** sono introdotti dalla formula lemma=;
- (iij) **cassazioni ed espunzioni** non sono state incluse nel testo del CT ma solo in msform e/o philform;
- (iiij) il **bastone di annotazione** è della forma "token\_lemma=lemma,HDF/HCF,MSF1,MSF2,MSF3,MSF4,MSF5".

21.1.5 LE CIFRE COMPLESSIVE DEL CT SECONDO "CWB-DESCRIBE-CORPUS". Viste le cifre generali del CT (§ 21.1.0) e l'architettura complessiva dei suoi attributi (§ 21.1.1), possiamo ora fornire i dati quantitativi analitici per ciascun tipo di posizione del CT. Per fare ciò il CWB dispone di un apposito comando.

21.1.5.1 IL "CWB-DESCRIBE-CORPUS". [MT] Una volta terminata con successo la procedura di preparazione del corpus, con la relativa installazione dello stesso in un'area della

memoria di massa appositamente creata allo scopo, è infatti possibile estrarre tutte le informazioni pertinenti la struttura del corpus in oggetto, operazione di primaria importanza non fosse che per finalità di controllo della coerenza strutturale del prodotto. Per ordinare al sistema di fornire tali informazioni, l'utente dovrà fare ricorso al comando del Corpus Work-Bench `cwb-describe-corpus` che, è bene ricordare, pur facendo parte della piattaforma di gestione CWB, risulta a tutti gli effetti estraneo al sistema di indagine CQP, e pertanto irraggiungibile anche dall'interfaccia web che abbiamo predisposto per il CT online (cfr. § 20.3).

In merito al funzionamento pratico, è importante ed utile segnalare che il comando in questione consente di ottenere i dati in due formati: uno di base ed uno più dettagliato.

21.1.5.2 IL "CWB-DESCRIBE-CORPUS" BASICO. [MT] Digitando il comando senza alcun elemento aggiuntivo, verranno restituiti unicamente i dati relativi al posizionamento del corpus all'interno dell'elaboratore, la dimensione del corpus espressa in termini di quantità complessiva di token e, infine, la quantità totale di attributi posizionali e strutturali, con il relativo elenco delle voci dichiarate in fase di *encoding*.

Nel caso dell'ultima installazione della corrente versione 1.8 che abbiamo fatto in locale in `bmanuel.org`, questa era la schermata che il sistema restituisce:

```

=====
Corpus: ant2009
=====

description:      CT18=ItalAnt2009, Apr 2009
registry file:    /usr/local/cwb-3.0/registry/ant2009
home directory:  /usr/etc/corpora/ant2009/
info file:       (none)
size (tokens):   257185

12 positional attributes:
  word          lemma          pos          kat
  typ           corr           genre        msform
  philform     mwlword        mwlkat       mwlnum

10 structural attributes:
  author        title          chapter      s
  line          page          par          type
  genr         mwl

0 alignment attributes:

```

Tav. 255: Descrizione basica del CT fornita da `cwb-describe-corpus`.

21.1.5.3 IL "CWB-DESCRIBE-CORPUS" ESPANSO. [MT] Per contro, specificando il parametro `-s` al comando di cui sopra, si otterrà l'espansione della sezione dedicata all'elenco delle voci appartenenti agli attributi strutturali e posizionali, con la relativa esplicitazione, per ogni entità presente, della quantità totale di `type` e `token`.

Nel caso dell'ultima installazione della corrente versione 1.8 che abbiamo fatto in locale questa è la schermata che il sistema, similmente a quella presentata in Tav. 255, restituisce con il comando `cwb-describe-corpus -s`: