

Corpus Taurinense: Sample Queries

Ulrich Heid

Draft input to the Web Presentation of CT
Version of 18/9/2000, printed on 26. Oktober 2000

Sample queries

The following sample queries give an overview of the types of information you can find in the *Corpus Taurinense* and the kinds of queries you can ask via the GUI.

Word forms, lemmas, word sequences

- *Word forms*: As a default, the standard word form is used in all queries. See query (13) in section 0.1.4 for cases where the manuscript and the philological form differ.

Example:

```
(1) Query:  "si";
      Results:
           per lui <si> manofesta lo dolzore e doglia con color trafigurato
```

```
(2) Query:  "cane";
      Results:
           Questi fu ÷e avarissimo come <cane> .
```

- *Words found by means of their lemma*: Inflected forms have the base form attached, and you can search for a base form and get all inflected cases.

Example:

```
(3) Query:  [lemma = "volere"];
      Results:
           parole <vorìa> conte avere ,
           e già non <volle> el meo detto primero c' Amor sua via a
           tutta gente aprisse
```

- *Clitics attached to a word form*: Clitics which are attached to a word form are marked with a special symbol.

The example below allows for "si" attached to a word form, written with one or two -s- (-ssi or -si): Note the explicit query for several clitics in (6).

```
(4) Query:  "÷.?si";
      Results:  -- [...] e va ÷vi <÷si> per una stretta via [...]
               -- e fa <÷ssi> da te , perché sa che se' fondamento
               ÷ll' altre .
```

- *Word sequences*: Several word forms in a row (sequence of expressions as in (1) may be searched for. Or a word form followed by any form of a lemma (see (3)), etc.

(5) Query: "si" "nasco.+";

Results:

```
-- E però è così terribile a vedere , perché ne ÷ll' idole che
adoravano li antichi <si nascondiano> i demoni , e facían
÷si a ÷lle genti adorare;
-- Il topo <si nascose> tra la farina e la gatta si mangiò la
crostata e , quand' ella aperse , il topo ne saltò fuori e
la gatta, perch' era satolla , non lo prese .
```

- *Clitic sequences*: As clitics are introduced by a special symbol (only if attached to a word form), the query below, which is a variant of a query for a word sequence gets you forms with clitic sequences. The symbol [] stands for an arbitrary word; it is followed by three clitics (i.e. three words, each starting with a clitic symbol):

(6) Query: [] "÷.+" "÷.+" "÷.+";

Results:

```
Ben non avemo noi da ÷' nostri compangni di Campangnia né di Fiandra
che ancora sieno de ÷l tutto diliveri , ma ongni die n' atendiamo
novelle che c osì ssia : <mandi ÷le ÷ci ÷ne> Iddio buone .
```

```
Il primo ke provò disse : Va , <porta ÷l ÷te ÷ne> : sì come tu ài
fatto il male , così te ne porta la pena , ké qua entro no ÷l
lasciera' tu .
```

Attributes of word forms

- *Word class*: the part of speech labelling of *Corpus Taurinense* is quite detailed (see the explanations in the *POS inventory* (link!)). For verbs, for example, infinitives are marked with a component "v." (for *verbs*) and "inf." (for *infinitives*). An example of a simple query for infinitives after *volere* is given below:

(7) Query: [lemma = "volere"] []+ [pos = ".*inf.*"] within s;

Results:

```
-- Di cupidità <vuol esser servito> ,
-- io <voglio perfettamente visitare>
-- parole <vorìa conte avere> ,
-- <Vuo' tu vedere> come le ricchezze e la gloria de
non vi <voglio eternalmente dannare>
```

The query allows for several items (of any kind) between *volere* and the infinitive.

- *Morphosyntax*: The *kat*-attribute allows for further morphosyntactic refinements of word class queries. The following query uses also the *POS*-feature to indicate the second person. The query explicitly looks for questions, which have the question mark as last element of the sentence:

(8) Query: [lemma = "volere" & pos = ".*ind.pr.*" & kat = ".*2,0,6.*"]
[]+ "?" within s;

Results:

```
E <vuo' promettere d' osservare i nostri ammonimenti ?>
<Vuo' tu , figliuolo , diventare nostro fedele ?>
```

Text type information

- *Text type*: We distinguish *verse*, *prosa* and *rubrica*. these can be queried via word forms, by means of the *typ*-feature; The following query checks all forms of the lemma *cuore* which are written with *-uo-*, and which appear in *verse*.

(9) Query: [lemma = "cuore" & word = ".+uo.+" & typ = "V"];
 Results:

13675: E però disse santo Ambruogio : << Chi ne dentro da ÷l su o <cuore> alberga Cristo , di smi delettamenti pasce l' anima sua >> .

14730: O dilettevole cena , quando Idio , cui tu ricevi ad albergo ne ÷l tuo <cuore> , quando per ę amore l' abbracce e lo stringi !

19613: E di questa prudenzia fa menzione Salamone dice : << Con ogni diligenza guarda il <cuor> tu vedi che , dicendo << guarda >> ,

The next query looks for the writing *core* in verse:

(10) Query: [lemma = "cuore" & word != ".+uo.+" & typ = "V"];
 Results:

442: Il fior d' amore è 'l primo nascimento de lo ch' è posto nascoso; la foglia?d è 'l disio c' ala monta; poi vèn lo frutto e guar' dà compimento di <core> è disioso, sì come il frutto che per sol s

541: Dolce potenza è quella de l' amore e ricca g spande tra gli amanti , che mette 'n ubrianza tant affanno , ca più sormonta in alegrezza un <core> u piacer d' amor , che non fan manti afanni l' om disavanzare in danno .

- *Genre* : The same way as for text types, we can also query for word forms in a particular genre, i.e. one of *doc(umentation)*, *did(actic writings)*, *stor* (for history), *nar(rative)* or *lir* (for lyrics). An example is given in the following pair of queries, where *ragione* has different senses, correlated with the presence in a documentary text or in another type:

(11) Query: [pos = ".*v.+ind.+"] []{1,4}[lemma = "ragione" & genre = "Doc"];
 Results:

<debbono rivedere la rasgione> de ÷' vecc hi camarlinghi .

po&(nemo&) che <de dare innanzi a sua ragone> , quatro perchamen einanzi .

in mezo giungno ne l' LXXXIIIIJ , lb~ DCCCLIIJ e s~ V e d~ IIIJ : <levamo da loro ragone> salda ove doveano

These queries also show simple ways of finding noun+verb-collocations.

(12) Query: `[pos = ".*v.+ind.+"] [1,4][lemma = "ragione" & genere != "Doc"];`
 Results:
`<è una sostanza senza ragione>: che , dacché non è buono
a [...]`

Corrections

We distinguish the form of a word which was found in the manuscript (*msform*), a standard form, and the philologically edited version, corrected and/or amended, of the manuscript form, *philform*.

Interesting cases are those (few) ones where the manuscript has a non-standard writing, i.e. where *msform* and *philform* differ. These cases are queried by means of an expression saying (roughly): ``for any form in the text ([]), check whether its (call it a) *msform*- differs from its *philform*-attribute". To express such a constraint, the item checked is made reference to by means of a variable name (here a):

(13) Query: `a:[] :: a.msform != a.philform`
 Results:
`<ricca/rica/ric&[c&]a>
<aggiate/agate/ag&[g&]iate>`

Über dieses Dokument ...

Corpus Taurinense: Sample Queries

This document was generated using the [LaTeX2HTML](#) translator Version 97.1 (release) (July 13th, 1997)

Copyright © 1993, 1994, 1995, 1996, 1997, [Nikos Drakos](#), Computer Based Learning Unit, University of Leeds.

The command line arguments were:

`latex2html -no_math -html_version 3.2,math -scalable_fonts -split 4 -link 6 queryinfo.tex.`

The translation was initiated by Wolfgang Lezius on 10/26/2000

Subsections

- [Sample queries](#)
 - [Word forms, lemmas, word sequences](#)
 - [Attributes of word forms](#)
 - [Text type information](#)
 - [Corrections](#)
- [Über dieses Dokument ...](#)

Next Up Previous

Wolfgang Lezius
 10/26/2000

Corpus Taurinense: Sample Queries

Ulrich Heid

Draft input to the Web Presentation of CT
Version of 18/9/2000, printed on 26. Oktober 2000

Sample queries

The following sample queries give an overview of the types of information you can find in the *Corpus Taurinense* and the kinds of queries you can ask via the GUI.

Word forms, lemmas, word sequences

- *Word forms*: As a default, the standard word form is used in all queries. See query (13) in section 0.1.4 for cases where the manuscript and the philological form differ.

Example:

(1) Query: "si";

Results:

per lui <si> manofesta lo dolzore e doglia con color trafigurato

(2) Query: "cane";

Results:

Questi fu ÷e avarissimo come <cane> .

- *Words found by means of their lemma*: Inflected forms have the base form attached, and you can search for a base form and get all inflected cases.

Example:

(3) Query: [lemma = "volere"];

Results:

parole <vorìa> conte avere ,

e già non <volle> el meo detto primero c' Amor sua via a
tutta gente aprisse

- *Clitics attached to a word form*: Clitics which are attached to a word form are marked with a special symbol.

The example below allows for "si" attached to a word form, written with one or two -s- (-ssi or -si): Note the explicit query for several clitics in (6).

(4) Query: "÷.?si";

Results: -- [...] e va ÷vi <÷si> per una stretta via [...]

-- e fa <÷ssi> da te , perché sa che se' fondamento
÷ll' altre .

- *Word sequences*: Several word forms in a row (sequence of expressions as in (1) may be searched for. Or a word form followed by any form of a lemma (see (3)), etc.

(5) Query: "si" "nasco.+";

Results:

```
-- E però è così terribile a vedere , perché ne ÷ll' idole che
adoravano li antichi <si nascondiano> i demoni , e facian
÷si a ÷lle genti adorare;
-- Il topo <si nascose> tra la farina e la gatta si mangiò la
crostata e , quand' ella aperse , il topo ne saltò fuori e
la gatta, perch' era satolla , non lo prese .
```

- *Clitic sequences*: As clitics are introduced by a special symbol (only if attached to a word form), the query below, which is a variant of a query for a word sequence gets you forms with clitic sequences. The symbol [] stands for an arbitrary word; it is followed by three clitics (i.e. three words, each starting with a clitic symbol):

(6) Query: [] "÷.+ " "÷.+ " "÷.+";

Results:

```
Ben non avemo noi da ÷' nostri compangni di Campangnia né di Fiandra
che ancora sieno de ÷l tutto diliveri , ma ongni die n' atendiamo
novelle che c osì ssia : <mandi ÷le ÷ci ÷ne> Iddio buone .
```

```
Il primo ke provò disse : Va , <porta ÷l ÷te ÷ne> : sì come tu ài
fatto il male , così te ne porta la pena , ké qua entro no ÷l
lasciera' tu .
```

Attributes of word forms

- *Word class*: the part of speech labelling of *Corpus Taurinense* is quite detailed (see the explanations in the *POS inventory* (link!)). For verbs, for example, infinitives are marked with a component "v." (for *verbs*) and "inf." (for *infinitives*). An example of a simple query for infinitives after *volere* is given below:

(7) Query: [lemma = "volere"] []+ [pos = ".*inf.*"] within s;

Results:

```
-- Di cupidità <vuol esser servito> ,
-- io <voglio perfettamente visitare>
-- parole <vorìa conte avere> ,
-- <Vuo' tu vedere> come le ricchezze e la gloria de
non vi <voglio eternalmente dannare>
```

The query allows for several items (of any kind) between *volere* and the infinitive.

- *Morphosyntax*: The *kat*-attribute allows for further morphosyntactic refinements of word class queries. The following query uses also the *POS*-feature to indicate the second person. The query explicitly looks for questions, which have the question mark as last element of the sentence:

(8) Query: [lemma = "volere" & pos = ".*ind.pr.*" & kat = ".*2,0,6.*"]
[]+ "\?" within s;

Results:

```
E <vuo' promettere d' osservare i nostri ammonimenti ?>
<Vuo' tu , figliuolo , diventare nostro fedele ?>
```

Text type information

- *Text type*: We distinguish *verse*, *prosa* and *rubrica*. these can be queried via word forms, by means of the *typ*-feature; The following

query checks all forms of the lemma *cuore* which are written with *-uo-*, and which appear in *verse*.

- (9) Query: [lemma = "cuore" & word = ".+uo.+" & typ = "V"];
Results:
- 13675: E però disse santo Ambruogio : << Chi ne dentro da ÷l su o <cuore> alberga Cristo , di smi delettamenti pasce l' anima sua >> .
- 14730: O dilettevole cena , quando Idio , cui tu ricevi ad albergo ne ÷l tuo <cuore> , quando per ę amore l' abbracce e lo stringi !
- 19613: E di questa prudenzia fa menzione Salamone dice : << Con ogni diligenza guarda il <cuor> tu vedi che , dicendo << guarda >> ,

The next query looks for the writing *core* in verse:

- (10) Query: [lemma = "cuore" & word != ".+uo.+" & typ = "V"];
Results:
- 442: Il fior d' amore è 'l primo nascimento de lo ch' è posto nascoso; la foglia?d è 'l disio c' ala monta; poi vèn lo frutto e guar' dà compimento di <core> è disioso, sì come il frutto che per sol s
- 541: Dolce potenza è quella de l' amore e ricca g spande tra gli amanti , che mette 'n ubrianza tant affanno , ca più sormonta in alegrezza un <core> u piacer d' amor , che non fan manti afanni l' om disavanzare in danno .

- *Genre* : The same way as for text types, we can also query for word forms in a particular genre, i.e. one of *doc(umentation)*, *did(actic writings)*, *stor* (for history), *nar(rative)* or *lir* (for lyrics). An example is given in the following pair of queries, where *ragione* has different senses, correlated with the presence in a documentary text or in another type:

- (11) Query: [pos = ".*v.+ind.+"] []{1,4}[lemma = "ragione" & genere = "Doc"];
Results:
- <debbono rivedere la rasgione> de ÷' vecc hi camarlinghi .
- po&(nemo&) che <de dare innanzi a sua ragone> , quatro perchamen einanzi .
- in mezo giungno ne l' LXXXIIIIJ , lb~ DCCCLIIJ e s~ V e d~ IIJ : <levamo da loro ragone> salda ove doveano
- These queries also show simple ways of finding noun+verb-collocations.

- (12) Query: [pos = ".*v.+ind.+"] []{1,4}[lemma = "ragione" & genere != "Doc"];
Results:
- <è una sustanzia senza ragione>: che , dacché non è buono a [...]

Corrections

We distinguish the form of a word which was found in the manuscript (*msform*), a standard form, and the philologically edited version, corrected and/or amended, of the manuscript form, *philform*.

Interesting cases are those (few) ones where the manuscript has a non-standard writing, i.e. where *msform* and *philform* differ. These cases are queried by means of an expression saying (roughly): ``for any form in the text ([]), check whether its (call it a) *msform*- differs from its *philform*-attribute". To express such a constraint, the item checked is made reference to by means of a variable name (here a):

(13) Query: `a:[] :: a.msform != a.philform`

Results:

`<ricca/rica/ric&[c&]a>`

`<aggiate/agate/ag&[g&]iate>`

Über dieses Dokument ...

Corpus Taurinense: Sample Queries

This document was generated using the [LaTeX2HTML](#) translator Version 97.1 (release) (July 13th, 1997)

Copyright © 1993, 1994, 1995, 1996, 1997, [Nikos Drakos](#), Computer Based Learning Unit, University of Leeds.

The command line arguments were:

`latex2html -no_math -html_version 3.2,math -scalable_fonts -split 4 -link 6 queryinfo.tex.`

The translation was initiated by Wolfgang Lezius on 10/26/2000

Subsections

- [Sample queries](#)
 - [Word forms, lemmas, word sequences](#)
 - [Attributes of word forms](#)
 - [Text type information](#)
 - [Corrections](#)
- [Über dieses Dokument ...](#)

[Next](#) [Up](#) [Previous](#)

Wolfgang Lezius

10/26/2000