

CONLL Format and Encoding Error Mark Up

Diana McCarthy and Siva Reddy

Lexical Computing Ltd

17th March 2011

Error Markup: Vertical File

wash NN wash-n
your PP\$ your-p
hands NNS hand-n
whenever WRB whenever-x
you PP you-p
sneeze VVP sneeze-v
into IN into-i
<err type='#A'>
it PP it-p
</err>
<g/>

Error Markup: Vertical File

```
wash NN wash-n
your PP$ your-p
hands NNS hand-n
whenever WRB whenever-x
you PP you-p
sneeze VVP sneeze-v
into IN into-i
<err type='#A'>
it PP it-p
</err>
<g/>
<corr type='#A'>
them PP them-p
</corr>
```

Vertical File

```
up RP up-x  
<err type='#WC'>  
looking VVG look-v  
horrible JJ horrible-j  
</err>  
<g/>  
so RB so-r  
watch VV watch-v
```

Vertical File

```
up RP up-x
<err type='#WC'>
looking VVG look-v
horrible JJ horrible-j
</err>
<g/>
<corr type='#WC'>
===NONE===
</corr>
so RB so-r
watch VV watch-v
```

Config File

```
STRUCTURE err {  
  ATTRIBUTE type {  
    LABEL "Error"  
  }  
  DISPLAYTAG 0  
  DISPLAYBEGIN "<%(type)>"  
  DISPLAYEND    "|"  
  NESTED 1  
}  
STRUCTURE corr {  
  ATTRIBUTE type {  
    LABEL "Correction"  
  }  
  DISPLAYTAG 0  
  DISPLAYBEGIN ""  
  DISPLAYEND    "</%(type)>"  
  NESTED 1  
}
```

```
SUBCORPATTRS "doc.domain|err.type|corr.type"
```

Demo or Slides

CUP's Learner Corpora: error mark up

Corpus: **Cambridge Learner Corpus Coded**

Hits: **1123**

Page of 57 [Next](#) | [Last](#)

561788_2	all the entertainments organized. My job	consisted	<#RT> in of </#R
1980180_2	Lee, Gun Hee in South Korea. The company	consists	of several divisions s
2252107_1	<#FN> courses courses </#S> course </FN><#AGV>	consist	consists </#AGV>
2252107_1	courses </#S> course </FN><#AGV> consist	consists	</#AGV> of 3 <#R
562128_2	national sales award. Our distribution network	consists	of many retail outlet
562204_2	Bank Security system <#MP> , </#MP>	consisting	of several alarms" <
562235_1	details </#S> about <#R> what is <#S> consited	consisted	</#S> of </#R> .
1707916_1	plan planning </#FV> the visit. The group	consists	<#MT> of </#MT>
1256636_2	As you know, our company's main service	consists	of organising busines
1707917_2	same. <#UP> Week-day Weekday </#UP> breakfast	consists	of bread, butter, jam
2248045_1	This process <#UV> is </#UV><#AGV>	consist	consists </#AGV>
2248045_1	process <#UV> is </#UV><#AGV> consist	consists	</#AGV><#RT> on
2248045_1	One third of <#MD> the </#MD> world	consists	<#RT> on of </#F
1803926_2	champion". Firstly, this programme <#AGV><#RV>	consist	consists </#AGV>
1803926_2	Firstly, this programme <#AGV><#RV> consist	consists	</#AGV> involves
2123549_2	Federation </#FN><#S> primarly primarily </#S>	consists	of <#MD> the </#
2251291_2	the afternoon between 4 pm and 7 pm. This	consists	<#MT> of </#MT>
562990_2	courtyard or in the fields. The neighbourhood	consisted	of several young fam
563023_2	or new discoveries. Yet another problem	consists	of becoming more af
563051_2	The activities <#UP> : </#UP><#RV>	consist	in </#RV> a sched

Page of 57 [Next](#) | [Last](#)

Error Queries

Concordance
Word List
Word Sketch
Thesaurus
Sketch-Diff

? Help on main menu

? Help on Expert Options

? Help on Error Query

Standard Query

Error Query

Expert options:

Context

Text Types

Switch menu position

Error Code



Incorrect word(s):

Corrected word(s):

[error codes](#)

Make Concordance

Clear All

Error Query: incorrect words

Concordance

Word List

Word Sketch

Thesaurus

Sketch-Diff

? Help on main menu

? Help on Expert Options

? Help on Error Query

Standard Query

Error Query

Expert options:

Context

Text Types

Switch menu position

Error Code



Incorrect word(s):

of

Corrected word(s):

[error codes](#)

Make Concordance

Clear All

Error Query: result

Corpus: **Cambridge Learner Corpus 2**

Hits: **6016**

Page of 301 [Next](#) | [Last](#)

1417824_2	They find parking difficult because <#R> of there is </#R> not enor
1072940_2	</#UT> your request for a reference <#RT> of for </#RT> Susan Gree
561788_2	<#RP> everyday every day </#RP> different sorts of <#CN> leisures leisure
561790_2	We sell our coils to different <#R> parts of the industry </#R> , e.
1097540_2	your letter. We accept your invitation <#RT> of to </#RT> the business
561802_1	stalls in Cooper's Park. In the leaflet <#RT> of about </#RT> the chari
561802_1	reliable <#AGN> newspapers newspaper </#AGN><#RT> of as </#RT> yours will ap
561802_1	enclosed </#TV> the financial figures <#RT> of for </#RT> the charity
1254006_1	</#RV> the courier and give me feedback <#RT> of on </#RT> the schedul
1070898_2	<#R> For all that and many other qualities of Ms. Green </#R> , whic
561814_2	sister smiling at her. At that point <#RT> of in </#RT> her life Mary
1095901_1	floor <#DT> oppositeside opposite </#DT><#R> of the </#R> sales depart
1095901_2	<#W> Thank you for inviting him, on behalf of John William, our sales m
1096334_1	costumers customers </#RN> in the <#W> region of Zurich Zurich region </#
1420223_1	cycling holiday to visit some places <#RT> of in </#RT> the country
759961_1	to 10.8 <#FN> billions billion </#FN><#UT> of </#UT> dollars, and the
759961_1	</#RV> 0.3 <#FN> billions billion </#FN><#UT> of </#UT><#FN> dollar
759961_1	with 10.8 <#FN> billions billion </#FN><#UT> of </#UT> dollars compar
561852_2	</#S> . This book is not only a guide <#RT> of to </#RT> the History o
561855_2	BECAUSE I GOT THIS RELAXING FEELING <#RT> OF FROM </#RT> FOCUSIN

Page of 301 [Next](#) | [Last](#)

Error Query: analysis by error code

Error	Freq	Rel [%]	
p/n #RT	2267	456.7	
p/n #R	1362	376.7	
p/n #UT	755	488.7	
p/n #W	460	244.4	
p/n #CE	303	1066.8	
p/n #U	226	565.9	
p/n #ID	134	349.6	
p/n #AS	92	306.2	
p/n #UN	87	221.2	
p/n #RV	52	11.1	-
p/n #RN	42	12.0	-
p/n #SX	36	30.9	-
p/n #DT	31	739.5	
p/n #DN	29	31.3	-
p/n #MT	24	13.9	-
p/n #RP	21	3.8	-
p/n #RJ	15	10.4	-
p/n #MP	13	2.3	-
p/n #RQ	12	64.1	
p/n #RY	6	5.4	-
p/n #L	6	9.7	-
p/n #DJ	6	5.9	-
p/n #IN	5	22.9	-
p/n #DY	5	10.4	-
p/n #HC	4	51.0	-

Error Query: querying codes

Concordance

Word List

Word Sketch

Thesaurus

Sketch-Diff

? Help on main menu

? Help on Expert Options

? Help on Error Query

Standard Query

Error Query

Expert options:

Context

Text Types

Switch menu position

Error Code :

Incorrect word(s):

Corrected word(s):

[error codes](#)

Make Concordance

Clear All

Error Query: result

Corpus: **Cambridge Learner Corpus Coded**

Hits: **420642**

Page of 21033 [Next](#) | [Last](#)

1417824_1	observed that spending on couriers <#RV>	is	has bee
1417824_2	<#RN> implementation </#RN> " I <#TV>	write	am writ
1417824_2	unsatisfactory </#RJ> <#RP> , we . We </#RP> need to <#RV>	give	provide
1417824_2	<#RP> , that . That </#RP> would help them <#MV>		do </#
1417824_2	more customers. Our opening hours <#AGV>	is	are </#
1417824_2	<#MA> it </#MA> difficult to <#RV>	search for	find </#
1417824_2	additional space for parking. Hope to <#RV>	achieve	do </#
561781_1	. <#UP> Stamp's Stamp </#UP> collecting <#TV>	is	has bee
561781_1	with it! After several days which I <#TV>	spend	spent <
561781_1	animals, sculptures, countries and so on, <#TV>	I've decided	I decide
561781_1	the </#MD> "Evening Club"? I <#TV>	also ask	would a
561781_1	ask </#TV> you <#UT> for </#UT> <#FV>	giving	to give
561781_2	. </#MP> Best Wishes Everybody <#AGV>	know	knows
561784_2	How much does a group of ten people <#FV>	costs	cost </#
1072940_2	is strong on teamwork and she always <#TV>	met	meets
1072940_2	the </#U> last year and she <#RV>	gained	made <
1072940_2	</#RY> . During the past 6 years, she <#MV>		has hac
1072940_2	days of sick <#CN> leaves leave </#CN> and <#UV>	applied	</#UV
561788_1	show that sport is still something <#TV>	looking	looked
561788_1	. In order to motivate young people <#FV>	playing	to play

Page of 21033 [Next](#) | [Last](#)