

## Corpus allemand

Annotation : Connexor

**Presse** **EMOCORP\_PRESSE\_DE** **121 076 999**

Description	Année	XML Id	Documents	Tokens
Frankfurter Allgemeine Zeitung	2002	faz02-cnx-prs-tei-utf8.xml	92 382	37 922 892
Frankfurter Rundschau	2008	fru08-cnx-prs-tei-utf8.xml	75 824	30 943 630
Der Tagesspiegel	2008	tasp08-cnx-prs-tei-utf8.xml	52 794	24 835 885
Hamburger Abendblatt	2008	hab08-cnx-prs-tei-utf8.xml	81 963	27 374 592

**Littérature** **EMOCORP\_LIT\_DE** **14 527 424**

Description	Année	XML Id	Documents	Tokens
Divers romans	1950-2010	litcorp_met_ge-cnx-prs-tei-utf8.xml	139	6 376 285
Divers romans	1950-2010	litmod_ge-cnx-prs-tei-utf8.xml	76	8 151 139

## Corpus anglais

Annotation : XIP

**Presse** **EMOCORP\_PRESSE\_EN** **99 594 161**

<b>Description</b>	<b>Année</b>	<b>XML Id</b>	<b>Documents</b>	<b>Tokens</b>
The Observer	2008	Guard08a-xrx-prs-tei.xml	60 230	31 503 083
Sunday Times	2008	Times08a-xrx-prs-tei.xml	75 285	37 970 290
The Independent On Sunday	2008	Indep08a-xrx-prs-tei.xml	31 986	21 055 509
Leicester Mercury	2008	LeiMer08-xrx-prs-tei.xml	31 478	9 065 279

**Littérature** **EMOCORP\_LIT\_EN** **37 931 114**

<b>Description</b>	<b>Année</b>	<b>XML Id</b>	<b>Documents</b>	<b>Tokens</b>
Divers romans	1950- 2010	litmod_en-cnx-prs-tei- utf8.xml	273	37 931 114

## Corpus espagnol

Annotation : Connexor

**Presse** **EMOCORP\_PRESSE\_ES** **127 357 935**

<b>Description</b>	<b>Année</b>	<b>XML Id</b>	<b>Documents</b>	<b>Tokens</b>
El País	2000	ep00-cnx-prs-tei-utf8.xml	58 562	30 676 134

El País	2008	ep08-cnx-prs-tei-utf8.xml	155 260	59 370 040
El Mundo	2008	em08-cnx-prs-tei-utf8.xml	58 266	37 311 761

**Littérature** **EMOCORP\_LIT\_ES** **28 727 735**

Description	Année	XML Id	Documents	Tokens
Divers romans	1950-2010	litmod_es-cnx-prs-tei-utf8.xml	294	28 727 735

## Corpus français

Annotation : Connexor

**Presse** **EMOCORP\_PRESSE\_FR** **120 874 622**

Description	Année	XML Id	Documents	Tokens
Le Monde	2007	lm07-cnx-prs-tei-utf8.xml	44 418	22 126 082
Le Monde	2008	lm08-cnx-prs-tei-utf8.xml	39 917	20 410 766
Le Figaro	2007	lfi07-cnx-prs-tei-utf8.xml	55 517	22 342 213
Le Figaro	2008	lfi08-cnx-prs-tei-utf8.xml	25 175	10 795 373
Libération	2007	libe07-cnx-prs-tei-utf8.xml	34 412	16 741 875
Ouest-France	2007	ouefr07-cnx-prs-tei-utf8.xml	67 183	16 325 284
Ouest-France	2008	ouefr08-cnx-prs-tei-utf8.xml	52 122	12 133 029

**Littérature** **EMOCORP\_LIT\_FR** **15 978 230**

Description	Année	XML Id	Documents	Tokens
Divers romans	1950-2010	litmod_fr-orig-cnx-prs-tei-utf8.xml	176	15 978 230

### Corpus russe

Annotation : SyntagRus/DeSR

**Presse** **EMOCORP\_PRESSE\_RU** **38 548 481**

Description	Année	XML Id	Documents	Tokens
АиФ Черноземье (АиФ на Дону / АиФ Калининград / АиФ Суперзвёзды / АиФ Тамбов / АиФ Челябинск / АиФ в Восточной Сибири / АиФ Томск / АиФ Черноземье / АиФ Тверь / АиФ на Енисее / АиФ Башкортостан / АиФ Я хочу всё узнать! / АиФ Белгород)	2004-2010	presse_ru-cnx-prs-tei-utf8.part1.xml	24 941	11 574 649
АиФ Черноземье (АиФ на Дону / АиФ Калининград / АиФ Суперзвёзды / АиФ Тамбов / АиФ Челябинск / АиФ в Восточной Сибири / АиФ Томск / АиФ	2004-2010	presse_ru-cnx-prs-tei-utf8.part2.xml	24 915	14 031 339

---

Черноземье / АйФ Тверь /  
АйФ на Енисее / АйФ  
Башкортостан / АйФ Я хочу  
всё узнать! / АйФ Белгород)

---

АйФ Черноземье (АйФ на  
Дону / АйФ Калининград /  
АйФ Суперзвёзды / АйФ  
Тамбов / АйФ Челябинск /  
АйФ в Восточной Сибири / 2004- presse\_ru-cnх-prs-tei-  
АйФ Томск / АйФ 2010 utf8.part3.xml 25 981 12 942 493  
Черноземье / АйФ Тверь /  
АйФ на Енисее / АйФ  
Башкортостан / АйФ Я хочу  
всё узнать! / АйФ Белгород)

---

**Littérature** **EMOCORP\_LIT\_RU** **99 256 442**

Description	Année	XML Id	Documents	Tokens
Divers romans	1950- 2010	litmod_ru-desr-prs-tei- utf8.part2.xml	600	34 475 215
Divers romans	1950- 2010	litmod_ru-desr-prs-tei- utf8.part3.xml	600	31 555 869
Divers romans	1950- 2010	litmod_ru-desr-prs-tei- utf8.part4.xml	600	33 225 358

---