THE LETRINT CORPORA

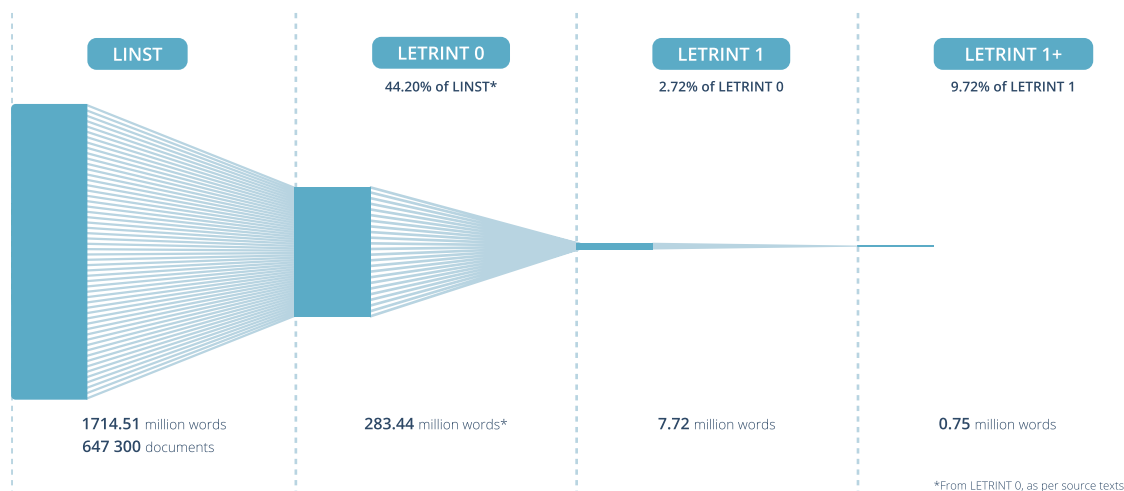| LINST	LETRINT 0	LETRINT 1	LETRINT 1+
	44.20% of LINST*	2.72% of LETRINT 0	9.72% of LETRINT 1
1714.51 million words	283.44 million words*	7.72 million words	0.75 million words
647 300 documents			

*From LETRINT 0, as per source texts.

The LETRINT corpora are four sets of trilingual textual datasets, including one comparable (LINST) and three parallel corpora (LETRINT 0, LETRINT 1 and LETRINT 1+). Their scope and features are determined by the goals of the LETRINT project. They comprise documents published in English, French and Spanish by the four main European Union institutions, the United Nations and its International Court of Justice, and the World Trade Organization in 2005, 2010 and 2015. This infographic explains the composition and the methodological details of each corpus. For further details, you can click on the "+" sign that appears at the top of each page. This resource provides a comprehensive overview of the corpora, but does not give access to the corpora themselves. A freely accessible corpus query interface, LETRINT-Q, has been developed for this purpose.

LINST
Three years of multilingual text production

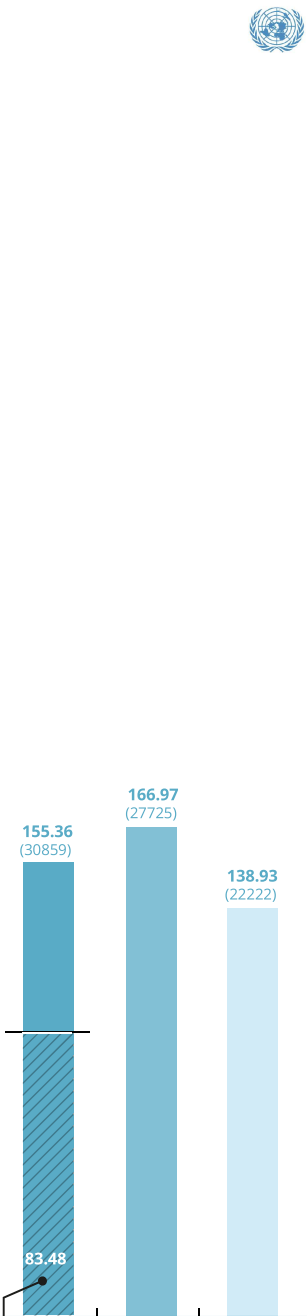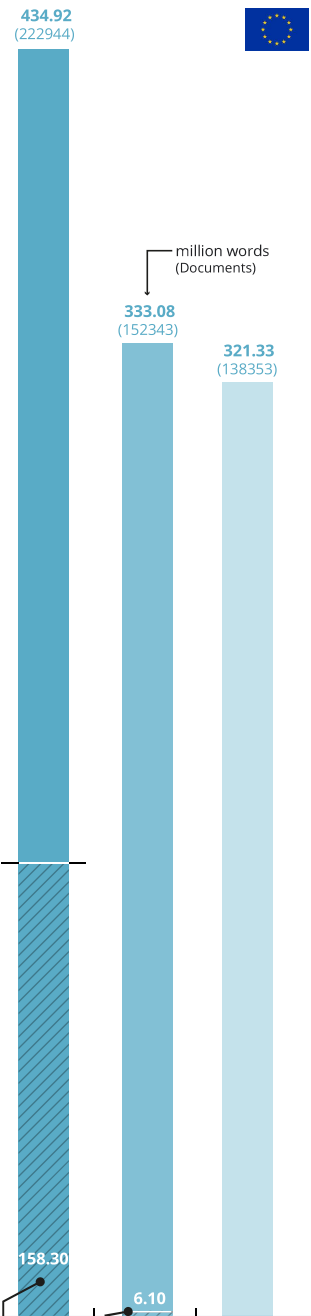All-inclusive compilation of publicly available texts in EN, FR or ES published in 2005, 2010 and 2015.

1714.51
million words
(647300 documents)

EN
FR
ES
Selected for LETRINT 0

EU: 1 089.33 million words (63.54%)
(513640 documents)

434.92
(222944)

million words
(Documents)

333.08
(152343)

321.33
(138353)

158.30

6.10

UN: 461.26 million words (26.90%)
(80806 documents)

155.36
(30859)

166.97
(27725)

138.93
(22222)

83.48

WTO: 163.92 million words (9.56%)
(52854 documents)

51.04
(17911)

55.66
(17458)

57.22
(17485)

35.55

LETRINT 0
Translation volumes per central legal function

Compilation of all documents translated from EN into FR and ES (or from FR into ES in the case of CJEU judgments), categorized by legal function.
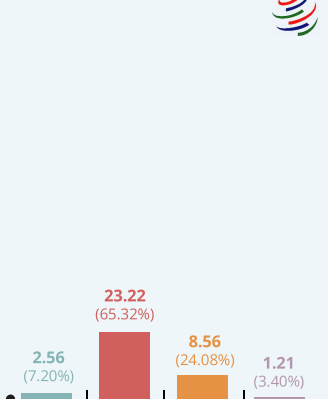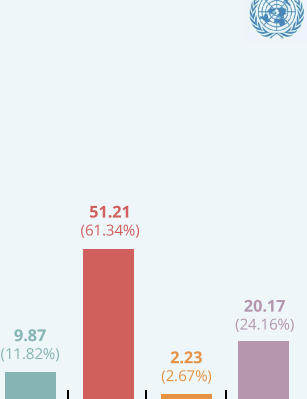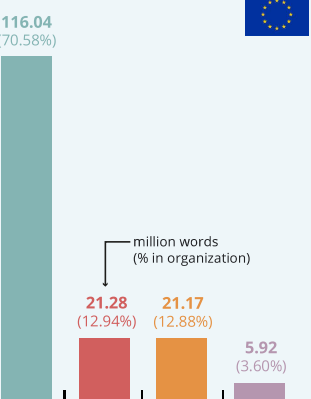
283.44
million words

1. Law- and policy- making
2. Monitoring
3. Adjudication
4. Administrative functions

* Source text word counts only.

EU: 164.41 million words

116.04
(70.58%)

million words
(% in organization)

21.28
(12.94%)

21.17
(12.88%)

5.92
(3.60%)

UN: 83.48 million words

9.87
(11.82%)

51.21
(61.34%)

2.23
(2.67%)

20.17
(24.16%)

WTO: 35.55 million words

2.56
(7.20%)

23.22
(65.32%)

8.56
(24.08%)

1.21
(3.40%)

44.20% of LINST
LETRINT 0

LETRINT 1
A comprehensive overview of institutional legal genres

Trilingual, parallel and part-of-speech annotated compilation of key legal genres in EN, ES and FR, selected through tailored stratified sampling.
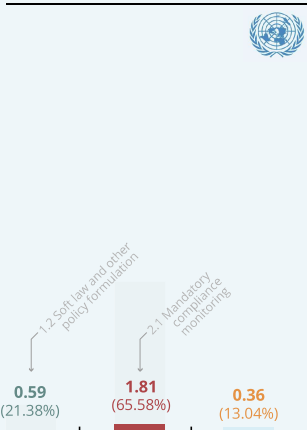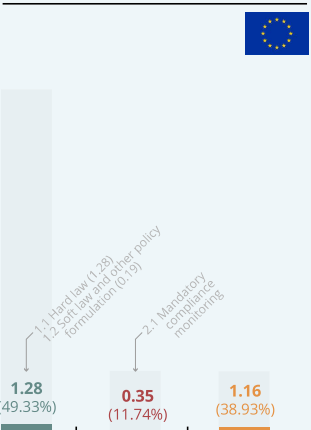
7.72
million words

1. Law- and policy- making
2. Compliance monitoring
3. Adjudication

EU: 2.98 million words

1.1 Hard law (1.28)
1.3 Soft law and other policy formulation (0.19)
2.1 Mandatory compliance monitoring

1.28
(49.33%)

0.35
(11.74%)

1.16
(38.93%)

UN: 2.76 million words

1.2 Soft law and other policy formulation
2.1 Mandatory compliance monitoring

0.59
(21.38%)

1.81
(65.58%)

0.36
(13.04%)

WTO: 1.98 million words

1.1 Hard law
2.1 Mandatory compliance monitoring

0.03
(1.52%)

1.03
(52.02%)

0.92
(46.46%)

2.72% of LETRINT 0
LETRINT 1

LETRINT 1+
An annotated selection of key institutional legal genres

Trilingual, parallel corpus of key legal genres compiled through down-sampling for systematic manual annotation of discourse features.
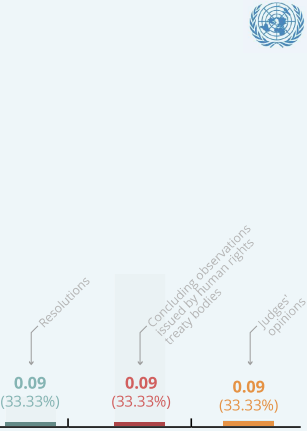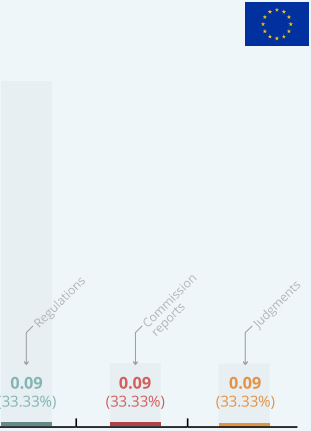
0.75
million words

1. Law-making
2. Compliance monitoring
3. Adjudication

EU: 0.27 million words

Regulations
Commission reports
Judgments

0.09
(33.33%)

0.09
(33.33%)

0.09
(33.33%)

UN: 0.27 million words

Resolutions
Concluding observations issued by human rights treaty bodies
Judges' opinions

0.09
(33.33%)

0.09
(33.33%)

0.09
(33.33%)

WTO: 0.21 million words

Binding decisions
Trade policy review reports
Dispute settlement reports

0.03
(14.30%)

0.09
(42.85%)

0.09
(42.85%)

9.72% of LETRINT 1+
LETRINT 1+