

Segmentation Rules

v 2

Directorate for Support and Technological Services for Translation

Euramis Pre-Translation Unit

Segmentation rules used in the European Parliament

What segmentation rules are used in the European Parliament?

The European Parliament uses specific segmentation rules, which are defined in the translation memories you receive from us. The stop characters are the full stop ('.') if immediately followed by a space and a word with capital initial letter, the colon (':'), the semicolon (';'), the exclamation mark ('!'), the question mark ('?'), the tabulator, the paragraph mark and the closing round bracket (')') if preceded by a maximum of three characters, such as 'aaa'.

Are the segmentation rules optional?

No, you must use our segmentation rules during translation. You are also requested to retain the segmentation of the original document wherever possible.

SDL STUDIO Memory templates

The table below sets out the segmentation rules used in Parliament' translation memory templates¹ with a description of the character strings before and after the segmentation in terms of regular expressions, examples of the patterns concerned, and example segments with the segmentation desired. The rules are presented as a main rule (including, where applicable, the options to be selected) followed by the exceptions to the rule, where appropriate.

		Name	Before	After	Patterns	Examples
1	Rule	Right Bracket Template_TM 1 Template_TM_el 1	<code>^\S][^\)]*[D][\p{Pe}\p{Pf}\p{Po}"]*</code>	<code>\s</code>	)	d) Ergebnisse der Analysen, Tests oder Diagnosen rechtzeitig liefern. (1a) Directive (EC)1234/2222 (taking into account) this very important matter + all the examples below in grey

¹ Template_TM (source language not EL)

- order of the rules: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- rules 11, 12 and 13 are not used

Template_TM_el (source language EL)

- order of the rules: 1, 11, 12, 2, 7, 5, 8, 13, 9, 3, 10
- rules 6 and 4 are not used

		Name	Before	After	Patterns	Examples
2	Rule	Full Stop Template_TM 2 Template_TM_el 4	\.+[p{Pe}p{Pf}p{Po}]"	\s	.	1. ¶ The Union reference laboratories and the national reference laboratories I. ¶ Révision de la directive sur les voyages à forfait
	- Check Abbreviations - Check Ordinal Followers - Include closing punctuation					
2-1	Exception	Lowercase-follower	.\w+\.	\s{p{L}}	Text. text	27. tammikuuta 2014
2-2	Exception	After Dot or Comma	[.,]\s?\w+\.	\s	, Text. , Text. . Text. . Text.	... text. Text. J.F. K. Der Handelsattaché der EU-Delegation erläutert, wie z. B. im Fall der Zulassung von Mobiltelefonanbietern eine Entscheidung von Fall zu Fall getroffen werde.
2-3	Exception	Bracket-follower	.\w+\.	not EL: \s[(\[] EL: \s[(\[]	Text. [text]	del regolamento (UE) n. [SR].

		Name	Before	After	Patterns	Examples
3	Rule	Tabulator Template_TM 3 Template_TM_el 10	.\t+	.	[Tab]	Θέμα[TAB] Ανεπαρκής δημοκρατική λογοδοσία της Τρόικας
		Name	Before	After	Patterns	Examples
4	Rule	Question mark Template_TM 4	[?]+\p{Pe}\p{Pf}\p{Po}"]*	\s	?	...Sanktionen nicht ausnutzt? Der Plutonium-Reaktor...
	■ Include closing punctuation					
4-1	Exception	Lowercase-follower	\?+\p{Pe}\p{Pf}\p{Po}"]*	\s[\p{Li}]	? text	Euramis, which lead to the development of Shout? and Twist.
		Name	Before	After	Patterns	Examples
5	Rule	Colon Template_TM 5 Template_TM_el 6	[:]+\p{Pe}\p{Pf}\p{Po}"]*	\s	:	ÜBERTRÄGER: Schweine, Pferde, Wildschweine und andere Wildtiere. από τις 9 Οκτωβρίου 2012: Μέλος της διακομματικής κοινοβουλευτικής ομάδας για τα άτομα με αναπηρία.
	■ Include closing punctuation					

		Name	Before	After	Patterns	Examples
6	Rule	Semicolon Template_TM 6	;+[p{Pe}\p{Pf}\p{Po}"]*	\s	;	...agreement on the MFF 2014-2020; notes that the negotiations revealed a wide gap between net contributors and net recipients
	■ Include closing punctuation					
		Name	Before	After	Patterns!	Examples
7	Rule	CharDotChar Template_TM 7 Template_TM_el 5	\w.\w	\s	XX	P.a Latvian text.
7-1	Exception	CharCharDotChar	.\w.\w	\s	Text.X	
		Name	Before	After	Patterns!	Examples
8	Rule	NumberDotChar Template_TM 8 Template_TM_el 7	\d+.\w	\s	9999.Text	1.a LV and HR units have their own way of numbering

8-1	Exception	CharNumberDotChar	.\d+\.\lw	\s	X9999. Text	... total cost of 100.000 €...
		Name	Before	After	Patterns	Examples
9	Rule	Exclamation mark Template_TM 9 Template_TM_el 9	[!]+\p{Pe}\p{Pf}\p{Po}"]*	\s	!	...Sanktionen nicht ausnutzt! Der Plutonium-Reaktor...
	Include closing punctuation					
9-1	Exception	Lowercase-follower	!+\p{Pe}\p{Pf}\p{Po}"]*	\s[\p{LI}]	! text	Euramis, which lead to the development of Shout! and Twist.
		Name	Before	After	Patterns	Examples
10	Rule	DotQuotation Template_TM 9 Template_TM_el 9	\.["]+	\s	."	The Commission and the European Parliament shall ensure that this hearing is organised at the European Parliament, if appropriate together with such other institutions and bodies of the Union as may wish to participate, and that the Commission is represented at an appropriate level." The complexity of the question required the involvement

						of three committees in the organisation of the hearing.
	Include closing punctuation					
		Name	Before	After	Patterns	Examples
11	Rule	Greek Question mark Template_TM_el 2	;+[\p{Pe}\p{Pf}\p{Po}]'	\s	;	Πώς εντάσσεται η παραπάνω κατάσταση στις ενταξιακές διαπραγματεύσεις της Τουρκίας στην ΕΕ; Εντάσσεται η ανάγκη αποκατάστασης των αδικιών και βιαιοτήτων εις βάρος των
	Include closing punctuation					
11-1	Exception	Lowercase-follower	;+[\p{Pe}\p{Pf}\p{Po}]'	\s[\p{Ll}]	; text	την περίοδο 2014-2020 ήταν το αποτέλεσμα; • συντηρητικών και
		Name	Before	After	Patterns	Examples
12	Rule	Greek Semicolon (Middle Dot u00B7) Template_TM_el 3	[\u00B7][\p{Pe}\p{Pf}\p{Po}]'	\s	•	προετοιμασίας αυτής της έκθεσης • επισημαίνει τα εμφανή

12-1	Exception	Number-Stop	(^\W)[0-9]+[\u00B7]	\s	99•	επιτροπής SURE, τον Ιούνιο 2011• φρονεί ότι η έκθεσ
		Name	Before	After	Patterns	Examples
13	Rule	Greek Semicolon (Ano Teleia u0387) Template_TM_el 8	[\u0387](\p{Pe}\p{Pf}\p{Po})"	\s	.	τονίζει ότι η συζήτηση σχετικά με τους πόρους για την πολιτική συνοχής πρέπει να βασίζεται σε ανάλυση των αναγκών και να λαμβάνει υπόψη την αξιολόγηση της εφαρμογής της πολιτικής στην προηγούμενη προγραμματική περίοδο• υπενθυμίζει ότι η πολιτική συνοχής είναι μια επενδυτική πολιτική υπέρ της ανάπτυξης που καταπολεμά την κρίση και επομένως θα αποκτήσει σημασία ως ενωσιακό μέσο για την ενίσχυση της βιώσιμης ανάπτυξης, τη δημιουργία θέσεων απασχόλησης, την τόνωση των δημοσίων επενδύσεων, τη μείωση των κοινωνικών ανισοτήτων και της φτώχειας, την τόνωση της ενεργειακής μετάβασης και την καταπολέμηση της κλιματικής αλλαγής.

\p{Pe} = any kind of closing bracket.

\p{Pf} = any kind of closing quote.

\p{Po} = any kind of punctuation character that is not a dash, bracket, quote or connector.

\s = Whitespace (including spaces)