

Worksheet 10.1 Parallel Texts

This resource is also on the slides but it's easier for the students to examine the data on their own worksheet. Solutions are at the end of this document and on Machine Translation slides.

10.1.1. An example of an English - Spanish Parallel Text

English	Spanish
Torres and associates.	Torres y asociados.
His clients are angry.	Sus clientes están enfadados.
The associates are also angry.	Los asociados también están enfadados.
The company has three groups.	La empresa tiene tres grupos.
The groups are in Europe.	Los grupos están en Europa.
The modern groups sell strong pharmaceuticals.	Los grupos modernos venden medicinas fuertes.
The groups do not sell aspirin.	Los grupos no venden aspirina.
The small groups are not modern.	Los grupos pequeños no son modernos.

Try to translate the following sentences into Spanish using the given parallel text above.

- a. Clients do not sell pharmaceuticals in Europe.

(Does the parallel text provide enough information?)

- b. The pharmaceuticals are very strong.

(Does the parallel text provide enough information?)



10.1.2 Another Parallel Text (English - Croatian)

English	Croatian
A mouse.	Miš.
The mouse.	Miš.
A cat.	Mačka.
The cat.	Mačka.
A cat chases a mouse.	Mačka juri miša.
A cat chases a mouse.	Miša juri mačka.
A mouse chases a cat.	Miš juri mačku.
A mouse chases a cat.	Mačku juri miš.
The cat is with the mouse.	Mačka je sa mišem.

<p>Notes about Croatian</p> <ul style="list-style-type: none"> • Fewer rules for word order (syntax) in Croatian • Richer morphology (many different word forms) • No articles 	<p>→ These differences between languages make learning from parallel texts more difficult.</p>
--	--

Translate:

- a. The mouse is with the cat.
-

10.1.3 Centauri and Arcturan - A parallel text when you don't know either language.

How might you translate between two languages you know nothing about? A parallel text.

Centauri	Arcturan
ok-voon ororok sprok .	at-voon bichat dat .
ok-drubel ok-voon anak plok sprok .	at-drubel at-voon pippat rrat dat .
erok sprok izok hihok ghirok .	totat dat arrat vat hilat .
ok-voon anak drok brok jok .	at-voon krat pippat sat lat .
wiwok farok izok stok .	totat jjat quat cat .
lalok sprok izok jok stok .	wat dat krat quat cat .
lalok farok ororok lalok sprok izok enemok .	wat jjat bichat wat dat vat eneat .
lalok brok anak plok nok .	iat lat pippat rrat nnat .
wiwok nok izok kantok ok-yurp .	totat nnat quat oloat at-yurp .
lalok mok nok yorok ghirok klok .	wat nnat gat mat bat hilat .
lalok nok crrrok hihok yorok zanzanok .	wat nnat arrat mat zanzanat .
lalok rarok nok izok hihok mok .	wat nnat forat arrat vat gat .

Translate this sentence in Centauri into Arcturan:

- a. farok crrrok hihok yorok klok kantok ok-yurp

Solutions to Worksheet 10.1 Parallel Texts

10.1.1. An example of an English - Spanish Parallel Text

- a. Yes! All necessary information can be found in the given parallel text and the translation is:
Clientes no venden medicinas en Europa.

 - b. The word “very” does not appear in the parallel text, so the text does not provide the information about the corresponding Spanish word
Las medicinas son ?? fuertes.
-

10.1.2 Another Parallel Text (English - Croatian)

- a. All words are occurring in the given parallel text but not all possible Croatian word *forms* are there

The translation is: Miš je sa mačkom.

but the form “mačkom” is not in the parallel text.

Machine translation needs as many forms as possible in data to work.

10.1.3 Centauri and Arcturan - A parallel text when you don't know either language.

- a. farok crrrok hihok yorok klok kantok ok-yurp

Arcturan words:

in the best order (according to the Arcturan part of the parallel text):

{jjat, arrat, mat, bat, oloat, at-yurp}