

# Natural Language Understanding, Generation, and Machine Translation

## Lecture 2: Machine Translation

---

Alexandra Birch

17 January 2024 (week 1)

School of Informatics  
University of Edinburgh  
[a.birch@ed.ac.uk](mailto:a.birch@ed.ac.uk)

Based on slides by Adam Lopez.

失败是成功之母。

失败是成功之母。

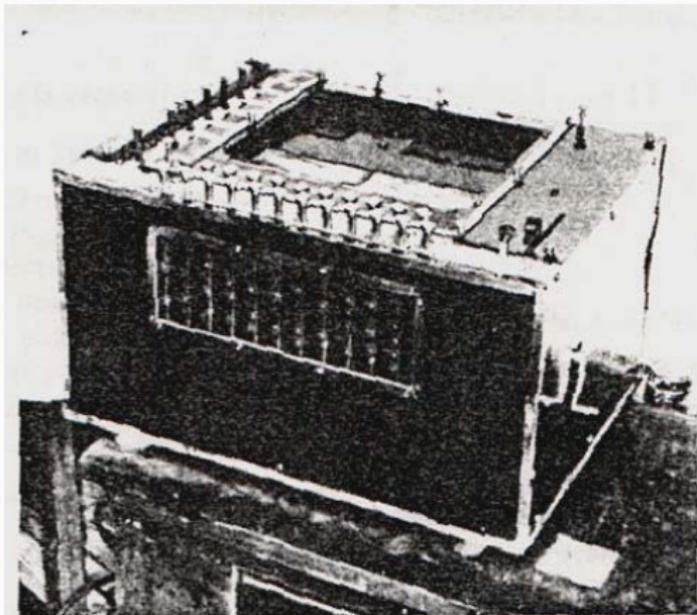
Failure is the mother of success.

# Historical Roots



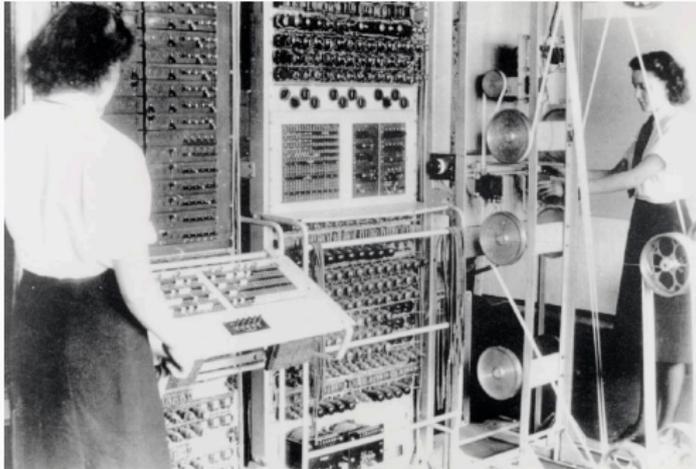
The Confusion of  
Tongues, Gustave  
Doré

## Earliest Machines



“Mechanical Brain”  
Translating Machine  
patented in 1933 in  
France

# First Computer



Dorothy Duboisson  
and Elise Booker  
(among world's first  
programmers) with  
Colossus (1944)

*Recognizing fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate.*

Warren Weaver (1947)

### Statistical machine translation live

Friday, April 28, 2006

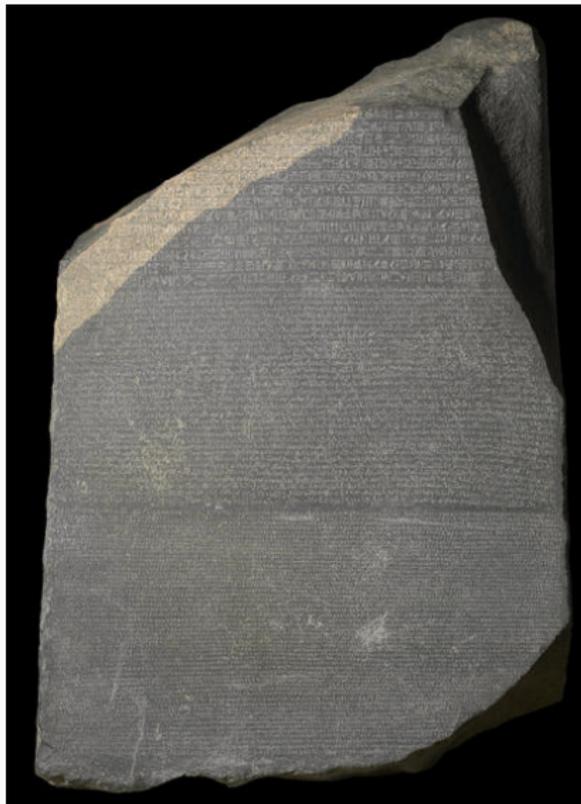
Posted by Franz Och, Research Scientist

Because we want to provide everyone with access to all the world's information, including information written in every language, one of the exciting projects at Google Research is machine translation. Most state-of-the-art commercial machine translation systems in use today have been developed using a rules-based approach and require a lot of work by linguists to define vocabularies and grammars.

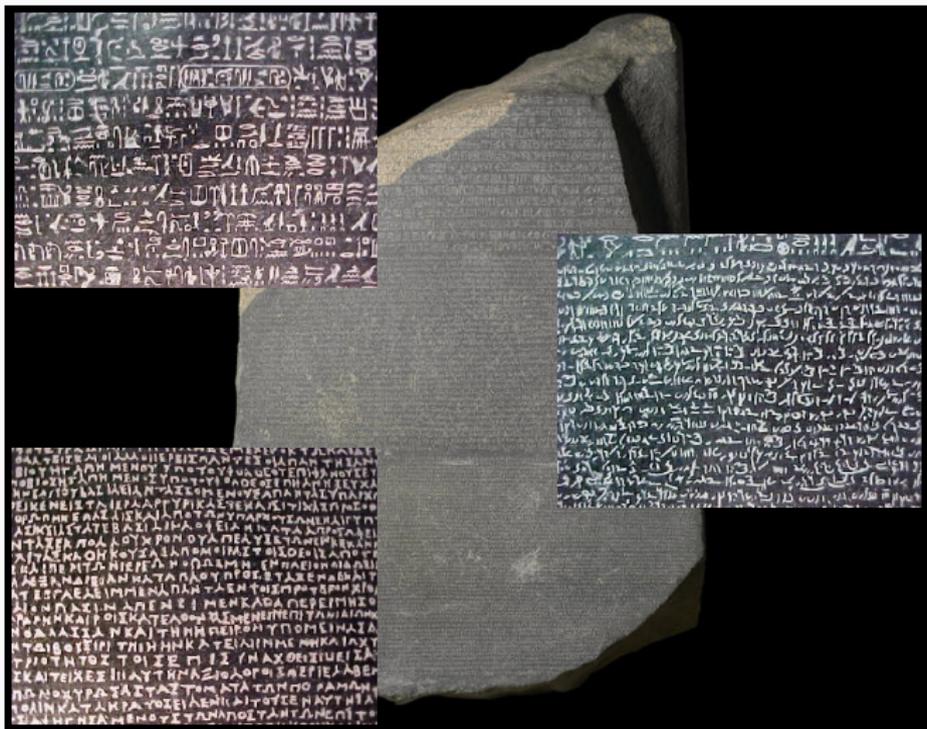
Several research systems, including ours, take a different approach: we feed the computer with billions of words of text, both monolingual text in the target language, and aligned text consisting of examples of human translations between the languages. We then apply statistical learning techniques to build a translation model. We have achieved [very good results](#) in research evaluations.

Now you can see the results for yourself. We recently launched an online version of our system for [Arabic-English](#) and [English-Arabic](#). Try it out!

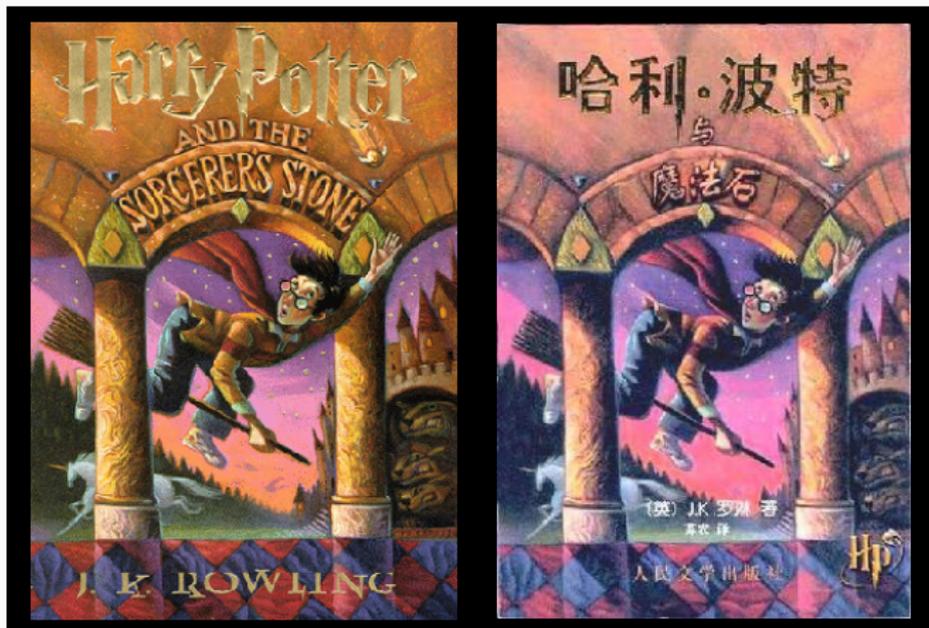
# Translation

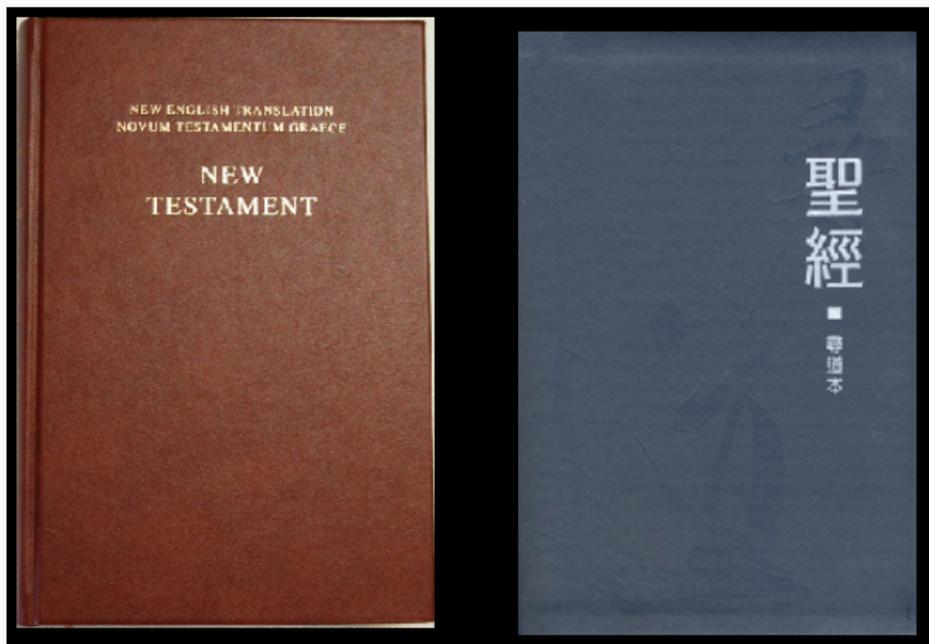


# Rosetta Stone



## Parallel Text





# Extraction of parallel data from web

Welcome to the United Nations العربية 中文 English Français Русский Español



**United Nations** | Climate Action

Home	<b>What Is Climate Change</b>	Raising Ambition >	The Science >	Actors, Actions, Solutions >
Act Now	UN and Climate Change >	Press Materials	Digital Library >	

## What Is Climate Change?

Climate change refers to long-term shifts in temperatures and weather patterns. These shifts may be natural, such as through variations in the solar cycle. But since the 1800s, human activities have been the main driver of climate change, primarily due to burning fossil fuels like coal, oil and gas.

Burning fossil fuels generates greenhouse gas emissions that act like a blanket wrapped around the Earth, trapping the sun's heat and raising temperatures.

Bienvenidos a las Naciones Unidas العربية 中文 English Français Русский Español



**Naciones Unidas** | Acción por el Clima

Portada	<b>¿Qué es el cambio climático</b>	Redoblar la ambición >	Ciencia >	Actores, acciones y soluciones >
¡Actúa ahora!	ONU y cambio climático >	Prensa (EN)	Biblioteca digital >	

## ¿Qué es el cambio climático?

El cambio climático se refiere a los cambios a largo plazo de las temperaturas y los patrones climáticos. Estos cambios pueden ser naturales, por ejemplo, a través de las variaciones del ciclo solar. Pero desde el siglo XIX, las actividades humanas han sido el principal motor del cambio climático, debido principalmente a la quema de combustibles fósiles como el carbón, el petróleo y el gas.

# Extraction of parallel data from web

Español

Welcome to the United Nations العربية 中文 **English** Français Русский Español



**United Nations** | Climate Action

Home | **What Is Climate Change** | Raising Ambition > | The Science > | Actors, Actions, Solutions > | Act Now | UN and Climate Change > | Press Materials | Digital Library >

## What Is Climate Change?

Climate change refers to long-term shifts in temperatures and weather patterns. These shifts may be natural, such as through variations in the solar cycle. But since the 1800s, human activities have been the main driver of climate change, primarily due to burning fossil fuels like coal, oil and gas.

Burning fossil fuels generates greenhouse gas emissions that act like a blanket wrapped around the Earth, trapping the sun's heat and raising temperatures.

English

Bienvenidos a las Naciones Unidas العربية 中文 **English** Français Русский **Español**



**Naciones Unidas** | Acción por el Clima

Portada | **Qué es el cambio climático** | Redoblar la ambición > | Ciencia > | Actores, acciones y soluciones > | ¡Actúa ahora! | ONU y cambio climático > | Prensa (EN) | Biblioteca digital >

## ¿Qué es el cambio climático?

El cambio climático se refiere a los cambios a largo plazo de las temperaturas y los patrones climáticos. Estos cambios pueden ser naturales, por ejemplo, a través de las variaciones del ciclo solar. Pero desde el siglo XIX, las actividades humanas han sido el principal motor del cambio climático, debido principalmente a la quema de combustibles fósiles como el carbón, el petróleo y el gas.

# Translation Exercise

# iat lat pippat eneat hilat oloat at-yurp .

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 .

# iat lat pippat eneat hilat oloat at-yurp .

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 .

## Translation Exercise

iat lat pippat eneat hilat oloat at-yurp .

arrat : hihok

at-drubel : ok-drubel

at-voon : ok-voon

at-yurp : ok-yurp

bat : klok

bichat : ororok

cat : stok

dat : sprok

eneat : enemok

forat : rarok

gat : mok

hilat : ghirok

iat : lalok

jjat : farok

krat : jok

lat : brok

mat : yorok

nnat : nok

oloat : kantok

pippat : anok

quat : izok

rrat : plok

sat : drok

totat : erok

totat : wiwok

vat : izok

wat : lalok

zanzanat : zanzanok

??? : crrrok

iat lat pippat eneat hilat oloat at-yurp .

arrat : hihok

at-drubel : ok-drubel

at-voon : ok-voon

at-yurp : ok-yurp

bat : klok

bichat : ororok

cat : stok

dat : sprok

eneat : enemok

forat : rarok

gat : mok

hilat : ghirok

iat : lalok

jjat : farok

krat : jok

lat : brok

mat : yorok

nnat : nok

oloat : kantok

pippat : anok

quat : izok

rrat : plok

sat : drok

totat : erok

totat : wiwok

vat : izok

wat : lalok

zanzanat : zanzanok

??? : crrrok



iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak enemok ghirook kantok ok-yurp .

erok sprok izok hihok ghirook .

totat dat arrat vat hilat .

ok-voon anak drok brok jok .

at-voon krat pippat sat lat .

lalok mok nok yorok ghirook klok .

wat nnat gat mat bat hilat .

lalok sprok izok jok stok .

wat dat krat quat cat .

lalok rarok nok izok hihok mok .

wat nnat forat arrat vat gat .

iat lat pippat eneat hilat oloat at-yurp .  
 \     /     \     /     \  
 lalok anak brok enemok kantok ghirook ok-yurp .

erok sprok izok hihok ghirook .  
 \     \     /     /  
 totat dat arrat vat hilat .

ok-voon anak drok brok jok .  
 \     /     /     /  
 at-voon krat pippat sat lat .

lalok mok nok yorok ghirook klok .  
 \     /     |     /  
 wat nnat gat mat bat hilat .

lalok sprok izok jok stok .  
 \     |     /     /  
 wat dat krat quat cat .

lalok rarok nok izok hihok mok .  
 \     /     /     /  
 wat nnat forat arrat vat gat .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak kantok ghirok enemok ok-yurp .

erok sprok izok hihok ghirok .

totat dat arrat vat hilat .

ok-voon anak drok brok jok .

at-voon krat pippat sat lat .

lalok mok nok yorok ghirok klok .

wat nnat gat mat bat hilat .

lalok sprok izok jok stok .

wat dat krat quat cat .

lalok rarok nok izok hihok mok .

wat nnat forat arrat vat gat .

iat lat pippat eneat hilat oloat at-yurp .  
 \ / \ / \ / \ / \ / \ /  
 lalok brok anak kantok ghirok enemok ok-yurp .

5040 possible orderings!

erok sprok izok hihok ghirok .  
 \ / \ / \ / \ /  
 totat dat arrat vat hilat .

ok-voon anak drok brok jok .  
 \ / \ / \ / \ /  
 at-voon krat pippat sat lat .

lalok mok nok yorok ghirok klok .  
 \ / \ / \ / \ /  
 wat nnat gat mat bat hilat .

lalok sprok izok jok stok .  
 \ / \ / \ / \ /  
 wat dat krat quat cat .

lalok rarok nok izok hihok mok .  
 \ / \ / \ / \ /  
 wat nnat forat arrat vat gat .



iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak enemok ghirook kantok ok-yurp .

ok-drubel anak ghirook farok .  
wiwok rarok nok zerok ghirook enemok .  
ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .  
lalok klok izok vok ok-drubel .  
ok-voon ororok sprok .  
ok-drubel ok-voon anak plok sprok .  
erok sprok izok hihok ghirook .  
ok-voon anak drok brok jok .  
wiwok farok izok stok .  
lalok sprok izok jok stok .  
lalok brok anak plok nok .  
lalok farok ororok lalok sprok izok enemok .  
wiwok nok izok kantok ok-yurp .  
lalok mok nok yorok ghirook klok .  
lalok nok crrrok hihok yorok zanzanok .  
lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak enemok ghirook kantok ok-yurp .

ok-drubel anak ghirook farok .  
wiwok rarok nok zerok ghirook enemok .  
ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .  
lalok klok izok vok ok-drubel .  
ok-voon ororok sprok .  
ok-drubel ok-voon anak plok sprok .  
erok sprok izok hihok ghirook .  
ok-voon anak drok brok jok .  
wiwok farok izok stok .  
lalok sprok izok jok stok .  
lalok brok anak plok nok .  
lalok farok ororok lalok sprok izok enemok .  
wiwok nok izok kantok ok-yurp .  
lalok mok nok yorok ghirook klok .  
lalok nok crrrok hihok yorok zanzanok .  
lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .

**lalok brok anak** enemok ghirok kantok ok-yurp .

ok-drubel anak ghirok farok .

wiwok rarok nok zerok ghirok enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirok .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

**lalok brok anak** plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirok klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .



iat lat pippat eneat hilat oloat at-yurp .  
lalok brok **anok enemok** ghirook kantok ok-yurp .

ok-drubel anak ghirook farok .

wiwok rarok nok zerok ghirook enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirook .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirook klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .



iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak enemok ghirokan tok ok-yurp .

ok-drubel anak ghirokan farok .

wiwok rarok nok zerok ghirokan enemok .

ok-drubel ziplok stok vok erok enemok kanton ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirokan .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kanton ok-yurp .

lalok mok nok yorok ghirokan klok .

lalok nok crrrok hihok yorok zanzanak .

lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak enemok ghirokan tok ok-yurp .

ok-drubel anak ghirokan farok .  
wiwok rarok nok zerok ghirokan enemok .  
ok-drubel ziplok stok vok erok enemok kanton ok-yurp zinok jok yorok klok .  
lalok klok izok vok ok-drubel .  
ok-voon ororok sprok .  
ok-drubel ok-voon anak plok sprok .  
erok sprok izok hihok ghirokan .  
ok-voon anak drok brok jok .  
wiwok farok izok stok .  
lalok sprok izok jok stok .  
lalok brok anak plok nok .  
lalok farok ororok lalok sprok izok enemok .  
wiwok nok izok kanton ok-yurp .  
lalok mok nok yorok ghirokan klok .  
lalok nok crrrok hihok yorok zanzanak .  
lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak **ghirok enemok** kantok ok-yurp .

ok-drubel anak ghirok farok .  
wiwok rarok nok zerok **ghirok enemok** .  
ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .  
lalok klok izok vok ok-drubel .  
ok-voon ororok sprok .  
ok-drubel ok-voon anak plok sprok .  
erok sprok izok hihok ghirok .  
ok-voon anak drok brok jok .  
wiwok farok izok stok .  
lalok sprok izok jok stok .  
lalok brok anak plok nok .  
lalok farok ororok lalok sprok izok enemok .  
wiwok nok izok kantok ok-yurp .  
lalok mok nok yorok ghirok klok .  
lalok nok crrrok hihok yorok zanzanok .  
lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak ghirok enemok kantok ok-yurp .

ok-drubel anak ghirok farok .  
wiwok rarok nok zerok ghirok enemok .  
ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .  
lalok klok izok vok ok-drubel .  
ok-voon ororok sprok .  
ok-drubel ok-voon anak plok sprok .  
erok sprok izok hihok ghirok .  
ok-voon anak drok brok jok .  
wiwok farok izok stok .  
lalok sprok izok jok stok .  
lalok brok anak plok nok .  
lalok farok ororok lalok sprok izok enemok .  
wiwok nok izok kantok ok-yurp .  
lalok mok nok yorok ghirok klok .  
lalok nok crrrok hihok yorok zanzanok .  
lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok **anok ghirook** enemok kantok ok-yurp .

ok-drubel anak ghirook farok .

wiwok rarok nok zerok ghirook enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirook .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirook klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok **anok ghirok** enemok kantok ok-yurp .

ok-drubel **anok ghirok** farok .  
wiwok rarok nok zerok ghirok enemok .  
ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .  
lalok klok izok vok ok-drubel .  
ok-voon ororok sprok .  
ok-drubel ok-voon anak plok sprok .  
erok sprok izok hihok ghirok .  
ok-voon anak drok brok jok .  
wiwok farok izok stok .  
lalok sprok izok jok stok .  
lalok brok anak plok nok .  
lalok farok ororok lalok sprok izok enemok .  
wiwok nok izok kantok ok-yurp .  
lalok mok nok yorok ghirok klok .  
lalok nok crrrok hihok yorok zanzanok .  
lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak ghirok enemok kantok ok-yurp .

ok-drubel anak ghirok farok .  
wiwok rarok nok zerok ghirok enemok .  
ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .  
lalok klok izok vok ok-drubel .  
ok-voon ororok sprok .  
ok-drubel ok-voon anak plok sprok .  
erok sprok izok hihok ghirok .  
ok-voon anak drok brok jok .  
wiwok farok izok stok .  
lalok sprok izok jok stok .  
lalok brok anak plok nok .  
lalok farok ororok lalok sprok izok enemok .  
wiwok nok izok kantok ok-yurp .  
lalok mok nok yorok ghirok klok .  
lalok nok crrrok hihok yorok zanzanok .  
lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak ghirook **enemok kantok** ok-yurp .

ok-drubel anak ghirook farok .

wiwok rarok nok zerok ghirook enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirook .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirook klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak ghirook **enemok kantok** ok-yurp .

ok-drubel anak ghirook farok .

wiwok rarok nok zerok ghirook enemok .

ok-drubel ziplok stok vok erok **enemok kantok** ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirook .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirook klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .

iat lat pippat eneat hilat oloat at-yurp .  
lalok brok anak ghirook enemok kantok ok-yurp .

ok-drubel anak ghirook farok .  
wiwok rarok nok zerok ghirook enemok .  
ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .  
lalok klok izok vok ok-drubel .  
ok-voon ororok sprok .  
ok-drubel ok-voon anak plok sprok .  
erok sprok izok hihok ghirook .  
ok-voon anak drok brok jok .  
wiwok farok izok stok .  
lalok sprok izok jok stok .  
lalok brok anak plok nok .  
lalok farok ororok lalok sprok izok enemok .  
wiwok nok izok kantok ok-yurp .  
lalok mok nok yorok ghirook klok .  
lalok nok crrrok hihok yorok zanzanok .  
lalok rarok nok izok hihok mok .



totat nnat forat arrat mat bat .

totat nnat forat arrat mat bat .

arrat : hihok

at-drubel : ok-drubel

at-voon : ok-voon

at-yurp : ok-yurp

bat : klok

bichat : ororok

cat : stok

dat : sprok

eneat : enemok

forat : rarok

gat : mok

hilat : ghirok

iat : lalok

jjat : farok

krat : jok

lat : brok

mat : yorok

nnat : nok

oloat : kantok

pippat : anok

quat : izok

rrat : plok

sat : drok

totat : erok

totat : wiwok

vat : izok

wat : lalok

zanzanat : zanzanok

??? : crrrok

totat nnat forat arrat mat bat .

arrat : hihok

at-drubel : ok-drubel

at-voon : ok-voon

at-yurp : ok-yurp

bat : klok

bichat : ororok

cat : stok

dat : sprok

eneat : enemok

forat : rarok

gat : mok

hilat : ghirok

iat : lalok

jjat : farok

krat : jok

lat : brok

mat : yorok

nnat : nok

oloat : kantok

pippat : anok

quat : izok

rrat : plok

sat : drok

totat : erok

totat : wiwok

vat : izok

wat : lalok

zanzanat : zanzanok

??? : crrrok

totat nnat forat arrat mat bat .  
erok | wiwok nok rarok hihok yorok klok .

arrat : hihok

at-drubel : ok-drubel

at-voon : ok-voon

at-yurp : ok-yurp

bat : klok

bichat : ororok

cat : stok

dat : sprok

eneat : enemok

forat : rarok

gat : mok

hilat : ghirok

iat : lalok

jjat : farok

krat : jok

lat : brok

mat : yorok

nnat : nok

oloat : kantok

pippat : anak

quat : izok

rrat : plok

sat : drok

totat : erok

totat : wiwok

vat : izok

wat : lalok

zanzanat : zanzanok

??? : crrrok

totat nnat forat arrat mat bat .  
erok | wiwok nok rarok hihok yorok klok .

ok-drubel anak ghirook farok .  
wiwok rarok nok zerok ghirook enemok .  
ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .  
lalok klok izok vok ok-drubel .  
ok-voon ororok sprok .  
ok-drubel ok-voon anak plok sprok .  
erok sprok izok hihok ghirook .  
ok-voon anak drok brok jok .  
wiwok farok izok stok .  
lalok sprok izok jok stok .  
lalok brok anak plok nok .  
lalok farok ororok lalok sprok izok enemok .  
wiwok nok izok kantok ok-yurp .  
lalok mok nok yorok ghirook klok .  
lalok nok crrrok hihok yorok zanzanok .  
lalok rarok nok izok hihok mok .

totat nnat forat arrat mat bat .  
erok | **wiwok nok** rarok hihok yorok klok .

ok-drubel anak ghirook farok .

wiwok rarok nok zerok ghirook enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirook .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

**wiwok nok** izok kantok ok-yurp .

lalok mok nok yorok ghirook klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .

totat nnat forat arrat mat bat .  
wiwok nok rarok hihok yorok klok .

ok-drubel anak ghirook farok .  
wiwok rarok nok zerok ghirook enemok .  
ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .  
lalok klok izok vok ok-drubel .  
ok-voon ororok sprok .  
ok-drubel ok-voon anak plok sprok .  
erok sprok izok hihok ghirook .  
ok-voon anak drok brok jok .  
wiwok farok izok stok .  
lalok sprok izok jok stok .  
lalok brok anak plok nok .  
lalok farok ororok lalok sprok izok enemok .  
wiwok nok izok kantok ok-yurp .  
lalok mok nok yorok ghirook klok .  
lalok nok crrrok hihok yorok zanzanok .  
lalok rarok nok izok hihok mok .

totat nnat forat arrat mat bat .  
wiwok rarok nok hihok yorok klok .

ok-drubel anak ghirook farok .

wiwok rarok nok zerok ghirook enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirook .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirook klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .

totat nnat forat arrat mat bat .  
wiwok rarok nok hihok yorok klok .

ok-drubel anak ghirook farok .

wiwok rarok nok zerok ghirook enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirook .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirook klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .

totat nnat forat arrat mat bat .  
wiwok rarok **nok hihok** yorok klok .

ok-drubel anak ghirok farok .

wiwok rarok nok zerok ghirok enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirok .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirok klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .

totat nnat forat arrat mat bat .  
wiwok rarok **nok hihok** yorok klok .

ok-drubel anak ghirok farok .

wiwok rarok nok zerok ghirok enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirok .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirok klok .

lalok **nok crrrok hihok yorok** zanzanok .

lalok rarok nok izok hihok mok .

totat nnat forat arrat mat bat .  
wiwok rarok **nok crrrok hihok yorok** klok .

ok-drubel anak ghirook farok .

wiwok rarok nok zerok ghirook enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirook .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirook klok .

lalok **nok crrrok hihok yorok** zanzanok .

lalok rarok nok izok hihok mok .

totat nnat forat arrat mat bat .  
wiwok rarok nok crrrok hihok yorok klok .

ok-drubel anak ghirook farok .

wiwok rarok nok zerok ghirook enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirook .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirook klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .

totat nnat forat arrat mat bat .  
wiwok rarok nok crrrok hihok **yorok klok** .

ok-drubel anak ghirok farok .

wiwok rarok nok zerok ghirok enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok **yorok klok** .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirok .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirok klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .

totat nnat forat arrat mat bat .  
wiwok rarok nok crrrok hihok yorok klok .

ok-drubel anak ghirok farok .

wiwok rarok nok zerok ghirok enemok .

ok-drubel ziplok stok vok erok enemok kantok ok-yurp zinok jok yorok klok .

lalok klok izok vok ok-drubel .

ok-voon ororok sprok .

ok-drubel ok-voon anak plok sprok .

erok sprok izok hihok ghirok .

ok-voon anak drok brok jok .

wiwok farok izok stok .

lalok sprok izok jok stok .

lalok brok anak plok nok .

lalok farok ororok lalok sprok izok enemok .

wiwok nok izok kantok ok-yurp .

lalok mok nok yorok ghirok klok .

lalok nok crrrok hihok yorok zanzanok .

lalok rarok nok izok hihok mok .

Is this exercise realistic?

In real natural language...

# Is this exercise realistic?

In real natural language...

...words are ambiguous

English: bank

French: banque

French: rive



# Is this exercise realistic?

In real natural language...

...words have complex *morphology*

<b>Tense</b>	<b>I</b>	<b>you</b>	<b>he, she, it</b>	<b>we</b>	<b>you</b>	<b>they</b>
<b>Present</b>	arrive	arrive	arrives	arrive	arrive	arrive
<b>Past</b>	arrived	arrived	arrived	arrived	arrived	arrived

# Is this exercise realistic?

In real natural language...

...words have complex *morphology*

	<b>singular</b>	<b>plural</b>
<b>nominative</b>	I	we
<b>oblique</b>	me	us
<b>possessive determiner</b>	my	our
<b>possessive pronoun</b>	mine	ours
<b>reflexive</b>	myself	ourselves

# Is this exercise realistic?

In real natural language...

...words have complex *morphology*

	<b>Singular</b>	<b>Plural</b>
<b>Nominativ</b>	das Haus	die Häuser
<b>Genitiv</b>	des Hauses	der Häuser
<b>Dativ</b>	dem Haus dem Hause	den Häusern
<b>Akkusativ</b>	das Haus	die Häuser

# Is this exercise realistic?

In real natural language...

...words have complex *morphology*

Finnish: ostoskeskuksessa

ostos#keskus+N+Sg+Loc:in

shopping#center+N+Sg+Loc:in

English: 'in the shopping center'

# Is this exercise realistic?

In real natural language...

...word order matters

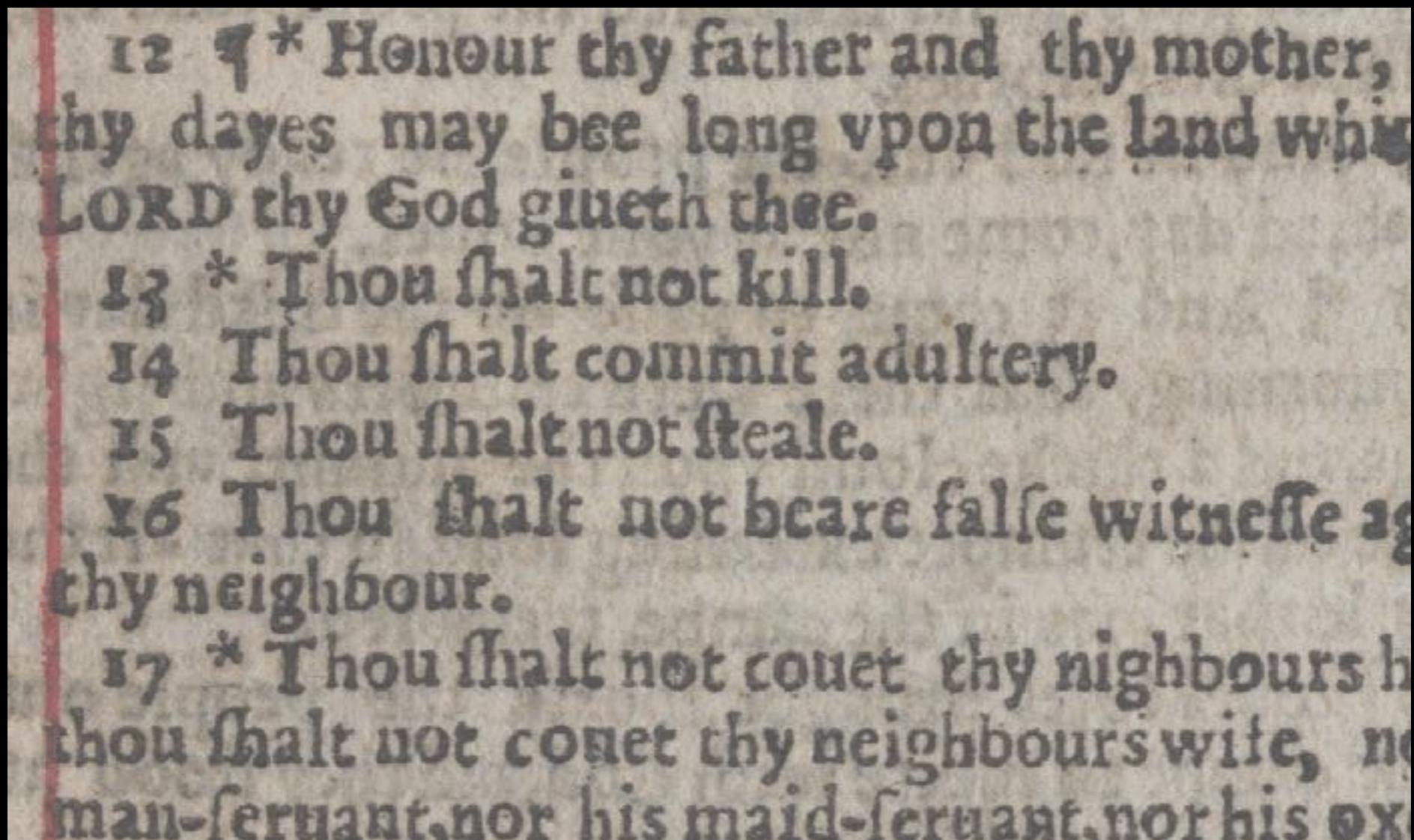
*Jaime loves Casey*

*Casey loves Jaime*

# Is this exercise realistic?

In real natural language...

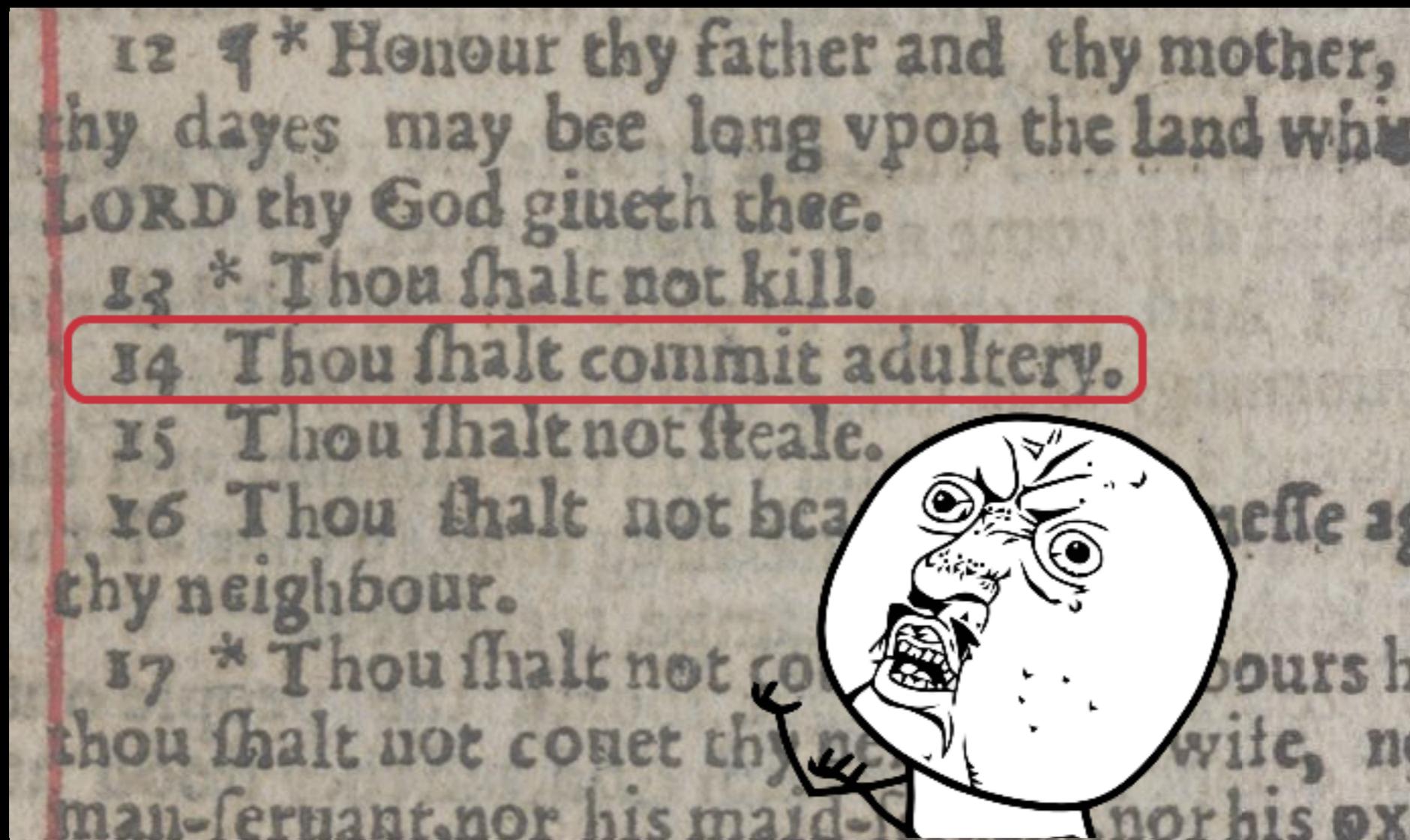
...every word matters



# Is this exercise realistic?

In real natural language...

...every word matters



# Is this exercise realistic?

In real natural language...

...sentences are longer and more complex.

*Greta Thunberg and fellow youth climate campaigners are demanding that global leaders immediately end the “madness” of huge ongoing investments in fossil fuel exploration and enormous subsidies for coal, oil and gas use.*

Is this exercise realistic?

iat lat pippat eneat hilat oloat at-yurp .  
 lalok brok anak ghirok enemok kantok ok-yurp .

ok-voon ororok sprok .  
 at-voon bichat dat .

lalok farok ororok lalok sprok izok enemok .  
 wat jjat bichat wat dat vat eneat .

ok-drubel ok-voon anak plok sprok .  
 at-drubel at-voon pippat rrat dat .

lalok brok anak plok nok .  
 iat lat pippat rrat nnat .

erok sprok izok hihok ghirok .  
 totat dat arrat vat hilat .

wiwok nok izok kantok ok-yurp .  
 totat nnat quat oloat at-yurp .

ok-voon anak drok brok jok .  
 at-voon krat pippat sat lat .

lalok mok nok yorok ghirok klok .  
 wat nnat gat mat bat hilat .

wiwok farok izok stok .  
 totat jjat quat cat .

lalok nok crrrok hihok yorok zanzanok .  
 wat nnat arrat mat zanzanat .

lalok sprok izok jok stok .  
 wat dat krat quat cat .

lalok rarok nok izok hihok mok .  
 wat nnat forat arrat vat gat .

# la empresa tiene enemigos fuertes en Europa .

Garcia and associates .

\ \ /  
Garcia y asociados .

Carlos Garcia has three associates .

\ \ | | /  
Carlos Garcia tiene tres asociados .

his associates are not strong .

| \ X /  
sus asociados no son fuertes .

Garcia has a company also .

\ / X /  
Garcia tambien tiene una empresa .

its clients are angry .

/ / | \  
sus clientes estan enfadados .

the associates are also angry .

/ / X \  
los asociados tambien estan enfadados .

the clients and the associates are enemies .

\ \ | / / / /  
los clientes y los asociados son enemigos .

the company has three groups .

\ | / / /  
la empresa tiene tres grupos .

its groups are in Europe .

/ | | \  
sus grupos estan en Europa .

the modern groups sell strong pharmaceuticals .

| X / \  
los grupos modernos venden medicinas fuertes .

the groups do not sell zanzanine .

| | / / /  
los grupos no venden zanzanina .

the small groups are not modern .

/ X X \  
los grupos pequenos no son modernos .

la empresa tiene enemigos fuertes en Europa .  
the company has strong enemies in Europe .

Garcia and associates .

Garcia y asociados .

Carlos Garcia has three associates .

Carlos Garcia tiene tres asociados .

his associates are not strong .

sus asociados no son fuertes .

Garcia has a company also .

Garcia tambien tiene una empresa .

its clients are angry .

sus clientes estan enfadados .

the associates are also angry .

los asociados tambien estan enfadados .

the clients and the associates are enemies .

los clientes y los asociados son enemigos .

the company has three groups .

la empresa tiene tres grupos .

its groups are in Europe .

sus grupos estan en Europa .

the modern groups sell strong pharmaceuticals .

los grupos modernos venden medicinas fuertes .

the groups do not sell zanzanine .

los grupos no venden zanzanina .

the small groups are not modern .

los grupos pequenos no son modernos .

## Key points to remember

- Very simple cues are very helpful in translation
  - Pairs of words that occur consecutively in the target language (bigrams)
  - Pairs of source and target words that occur frequently occur together in translations
- For future lectures:
  - How would you model these cues?
  - What are the ways in which this could fail?

Conditional Language Modeling with n-grams