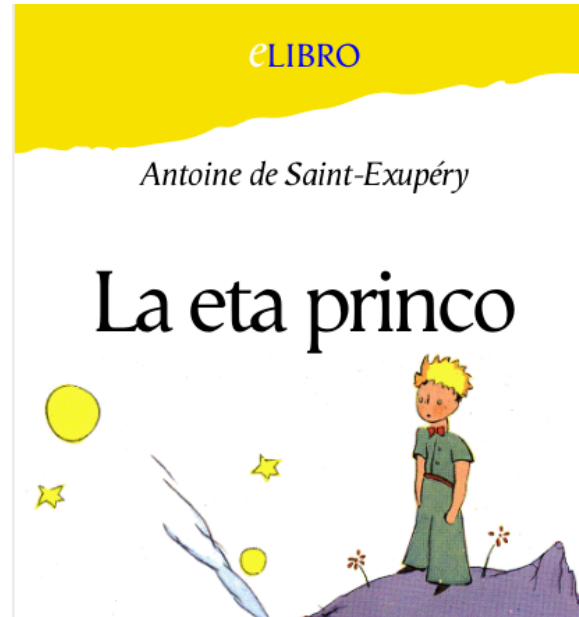


Statistical comparison of translations of "The Little Prince" by Antoine de Saint-Exupéry in different languages

Translated from Esperanto
by Uli Ender
and
Google Translate

Contact:
uli.ender@esperanto.de



Why this book?

The book has been translated into more than 500 languages.
It is formulated in simple language, understandable even to children.
It is very suitable for comparing and learning foreign languages.

01	english / angla	10	greek / greka	19	latvian / latva
02	bulgarian / bulgara	11	hebrew / hebrea	20	persian / persa
03	czech / čeňa	12	indian / hindia	21	polish / pola
04	chinese / čina	13	spanish / hispana	22	russian / rusa
05	Esperanto	14	hungarian / hungara	23	romanian / rumana
06	estonian / estona	15	italian / itala	24	serbian / serba
07	basque / eŭska	16	japanese / japana	25	thaj / taja
08	french / franca	17	georgian / kartvela	26	turkish / turka
09	german / germana	18	catalan / kataluna	27	vietnamese / vietnama

Preparing the texts

using copy/past from html-file using pdftotext for pdf-files

- problem of wrapping words/pages
- converting capital letters to lower case letters
- extracting punctuation

Analyzing the texts

using simple tcl-scripts

- a) number of words / various words
- b) number of letters / various letters
- c) number of syllables / various syllables
- d) number of word roots / various word roots (eg. Esperanto)

Preparing the texts using pdftotext

LA ETA PRINCO

eLIBRO

Dediĉo de la aŭtoro al Léon Werth

Pardonu infanoj, ke mi dediĉis ĉi tiun libron al "grandulo". Mi havas seriozan senkulpigon: tiu grandulo estas la plej bona amiko, kiun mi havas en la mondo. Mi havas alian senkulpigon: tiu grandulo kapablas ĉion kompreni, eĉ porinfanajn librojn. Mi havas trian senkulpigon: tiu grandulo loĝas en Fran-

.....

"Tiu ĉi estas tro maljuna. Mi volas ŝafeton, kiu longe vivu." Tiam mi senpacienciĝis kaj, ĉar mi volis senprokraste komen-

8

LA ETA PRINCO

eLIBRO

ci la malmuntadon de mia motoro, mi skizaĉis ĉi tiun lastan desegnon.

eLIBRO

LA ETA PRINCO

Dediĉo de la aŭtoro al Léon Werth

P

ardonu infanoj, ke mi dediĉis ĉi tiun libron al "grandulo". Mi havas seriozan senkulpigon: tiu grandulo estas la plej bona amiko, kiun mi havas en la mondo. Mi havas alian senkulpigon: tiu grandulo kapablas ĉion kompreni, eĉ porinfanajn librojn. Mi havas trian senkulpigon: tiu grandulo loĝas en Francio,

.....

"Tiu ĉi estas tro maljuna. Mi volas ĵafeton, kiu longe vivu." Tiam mi senpacienciĝis kaj, ĉar mi volis senprokraste komen-

eLIBRO

LA ETA PRINCO

ci la malmuntadon de mia motoro, mi skizaĉis ĉi tiun lastan desegnon.

नन्हा राजकुमार

सैंतेक्जूपेरी

अनुवाद: लाल बहादुर वर्मा

नन्हा राजकुमार (मूल फ्रेंच से अनुवाद)

uUgk jktdqekj
ISarsDtwisjh vuqokn% yky cgknqj
oekZ uUgk jktdqekj (ewy Úsap ls
vuqokn)

After additional conversions:



Pardonu infanoj, ke mi dediĉis ĉi tiun libron al „grandulo“.

Analyzing the texts:

If I have told you so many details about asteroid B 612 and if I have revealed its number, it is precisely for adults who love numbers.

a) words / various words

```
set t {Se vi ho raccontato tanti particolari sull'asteroide B 612 e se vi ho rivelato il suo numero,e proprio per i grandi che amano le cifre.}
```

omit punctuation & transform to lower size letters:

```
regsub -all {[.,;?!'']} $t " " t  
set t [string tolower $t]
```

sull'asteroide == su la asteroide !!!

→ se vi ho raccontato tanti particolari sull'asteroide b 612 e se vi ho rivelato il suo numero e proprio per i grandi che amano le cifre

count words & various words cont.

```
set n 0
set l [llength $t]
.t insert end "number of words $\n"
While {$n < $l} {
set Wort [lindex $t $n]
incr word($Wort)
incr n
}
set z [array get word]
set l [llength $z]
.t insert end "various words [expr $l/2]"
set n 0
set zz " "
while {$n < $l} {
set a [lindex $z $n]
incr n
set b [lindex $z $n]
set zz "$zz {$a $b}"
incr n
}
set zz [lsort -ascii -index 0 $zz]
set yy [lsort -integer -decreasing -index 1 $zz]

# finally print a table of words and number of
# occurance
```

number of words 27
various words 23

Word	number
=====	=====
e	2
ho	2
se	2
vi	2
612	1
amano	1
asteroide	1
b	1
che	1
cifre	1
grandi	1
...	
sull	1
suo	1
tanti	1

b) count characters & various characters

```
set n 0
.t insert end "\n\nAnalysis\n-----\n"
##### I is the number of characters #####
set I [string length $t]
.t insert end "number of characters $I\n"

##### find out how often characters occur #####
while {$n < $I} {
  set zeichen [string index $t $n]
  if {$zeichen == " "} {set zeichen "spaces"}
  incr letter($zeichen)
  incr n
}
set z [array get letter]
set I [llength $z]
.t insert end "various characters [expr $I/2]\n"
set n 0
set zz ""
while {$n < $I} {
  set a [lindex $z $n]
  incr n
  set b [lindex $z $n]
  set zz "$zz {$a $b}"
  incr n
}
...
```

Analysis

number of characters 136
various characters 27

character	number	utf-8-code
=====	=====	=====
spaces	25	\u0020
e	12	\u0065
i	12	\u0069
o	12	\u006f
r	11	\u0072
a	10	\u0061
...		
f	1	\u0066
6	1	\u0036
g	1	\u0067
'	1	\u0027
,	1	\u002c
.	1	\u002e

c) calculation number of syllables containing diphthongs and vowels

```
.t insert end \n\n$t\n
```

```
set n1 [regsub -all "oi" $t "xx" t]
```

```
.t insert end "\n$n1 x oi\n"
```

```
set n2 [regsub -all "io" $t "xx" t]
```

```
.t insert end "$n2 x io\n"
```

```
...
```

```
.t insert end "sum [expr ($n1 + $n2 + $n3 + $n4 + $n5 + $n6 + $n7)] syllables\n"
```

Se vi ho raccontato tanti . . .

1 x oi

1 x io

10 x a

12 x e

10 x i

10 x o

3 x u

sum 47 syllables

d) calculation of word roots in Esperanto

In Esperanto the meaning of a word is composed by the meanings of the root-words it consists of (The algorithm works as follows)

word	rest			

plenkreskuloj				
plenkreskulo	j			
plenkreskul	oj			
plenkresku	loj			
plenkresk	uloj			
plenkres	kuloj			
plenkre	skuloj			
plenkr	eskuloj			
plenk	reskuloj			
plen	kreskuloj	word root "plen"	found	meaning: completely
plen'kreskuloj				
plen'kreskulo	j			
plen'kreskul	oj			
plen'kresku	loj			
plen'kresk	uloj	word root "kresk"	found	meaning: grown-up
plen'kresk'uloj				
plen'kresk'ulo	j			
plen'kresk'ul	oj	word root "ul"	found	meaning: person
plen'kresk'ul'oj				
plen'kresk'ul'o	j	word root "o"	found	meaning: substantive
plen'kresk'ul'o'j		word root "j"	found	meaning: plural
- - > plenkreskuloj == adults !				

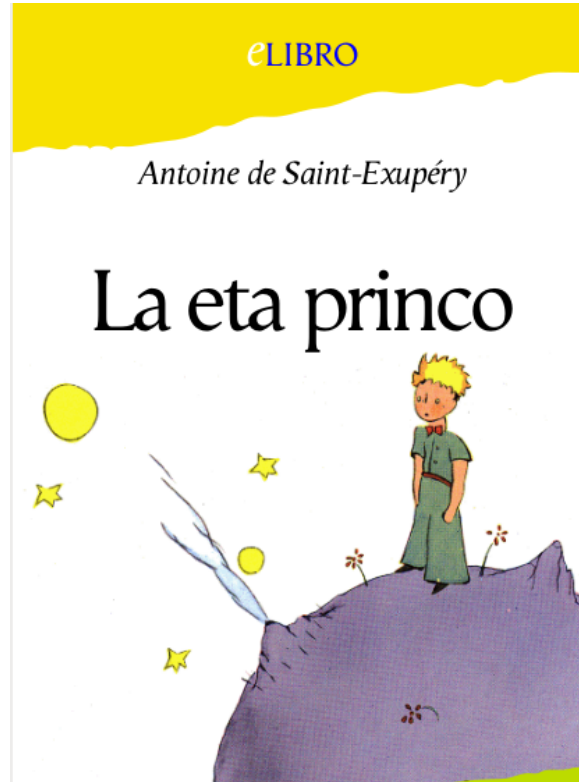
text
=====
se mi rakontis al vi tiom da
...

list of root-words:
=====
a' al am' as' asteroid' da detal' e'
est' ĝi' ĝust' is' j' kaŝ' kaj kiu' kresk'
la mal' mi n' nombr' numer' o' por
plen' pri rakont' se ŝip' tio tiom ul'
vapor' vi

words divided into root-words
=====
se mi rakont'is al vi tiom da
...

Statistical comparison of translations of "The Little Prince" by Antoine de Saint-Exupéry in different languages

Translated from Esperanto
by Uli Ender



A little about the book “The Little Prince”:

A pilot has a breakdown in the desert. When he is alone repairing the engine, he meets the little prince.

He comes from the asteroid B612 and tells about his life, his love for a flower, about dangerous baobabs and about meetings with strange people, such as a drunkard, a lamplighter, a switchman, while visiting other asteroids and the earth.

He philosophizes about friendship, love, about the different characters of adults and their meaning of life.

Famous are the phrases:

One sees well only with one's heart.

The essence is invisible to the eyes.

The book has been translated into more than 500 languages.

It is formulated in simple language, understandable even to children.

It is very suitable for comparing and learning foreign languages.

01	english / angla	10	greek / greka	19	latvian / latva
02	bulgarian / bulgara	11	hebrew / hebrea	20	persian / persa
03	czech / čeňa	12	indian / hindia	21	polish / pola
04	chinese / čina	13	spanish / hispana	22	russian / rusa
05	Esperanto	14	hungarian / hungara	23	romanian / rumana
06	estonian / estona	15	italian / itala	24	serbian / serba
07	basque / eũska	16	japanese / japana	25	thaj / taja
08	french / franca	17	georgian / kartvela	26	turkish / turka
09	german / germana	18	catalan / kataluna	27	vietnamese / vietnama

Differences from the two originals: 1943

1946



Antoine de Saint-Exupéry in Canada 1943



"I saw the sunset forty-four times!"



Un jour, j'ai vu le soleil se coucher quarante-trois fois!

"Unu tagon mi vidis la sunon subiri kvardek tri fojojn!"

My check: All translations contain the dedication to Leon Werth and the 27 chapters!
Catalan text : largely in Spanish – Hungarian text : one chapter in Czech

My check: All translations contain the dedication to Leon Werth and the 27 chapters!

The same sentence has been translated in different ways:

The french original from 1946 :

Si je vous ai raconté ces détails sur l'astéroïde B 612 et si je vous ai confié son numéro, c'est à cause des grandes personnes. Les grandes personnes aiment les chiffres.

The translation into Esperanto by Pierre Delaire (Canada):

Se mi rakontis al vi tiujn detalojn pri la asteroido B-612 kaj konfidis ĝian numeron, estas nur pro la granduloj. La granduloj frandas ciferojn.

The 1943 english translation probably by Katherine Woods (USA) shortened the two sentences.

If I have told you these details about the asteroid, and made a note of its number for you, it is on account of the grown-ups and their ways.

In the German translation by Mr. and Mrs. Leitgeb (Switzerland), the words "asteroid" and "detail" are avoided. They probably thought that german children would not understand these foreign words!

Wenn ich euch dieses nebensächliche Drum und Dran über den Planeten B 612 erzähle und euch sogar seine Nummer anvertraue, so geschieht das der großen Leute wegen. Die großen Leute haben eine Vorliebe für Zahlen.

To write a language, you must learn the various letters, ideograms (Chinese, Japanese), numbers, and punctuation marks.

For example, a period at the end of a sentence is a small circle in Chinese and Japanese 。 and in Hindi the dash | !

Chinese (六 一 二) and Thai (๖๑๒) languages, for example, use different symbols for numbers than European languages.

Chinese, Japanese and Thai languages do not differentiate between individual words

Ex.:

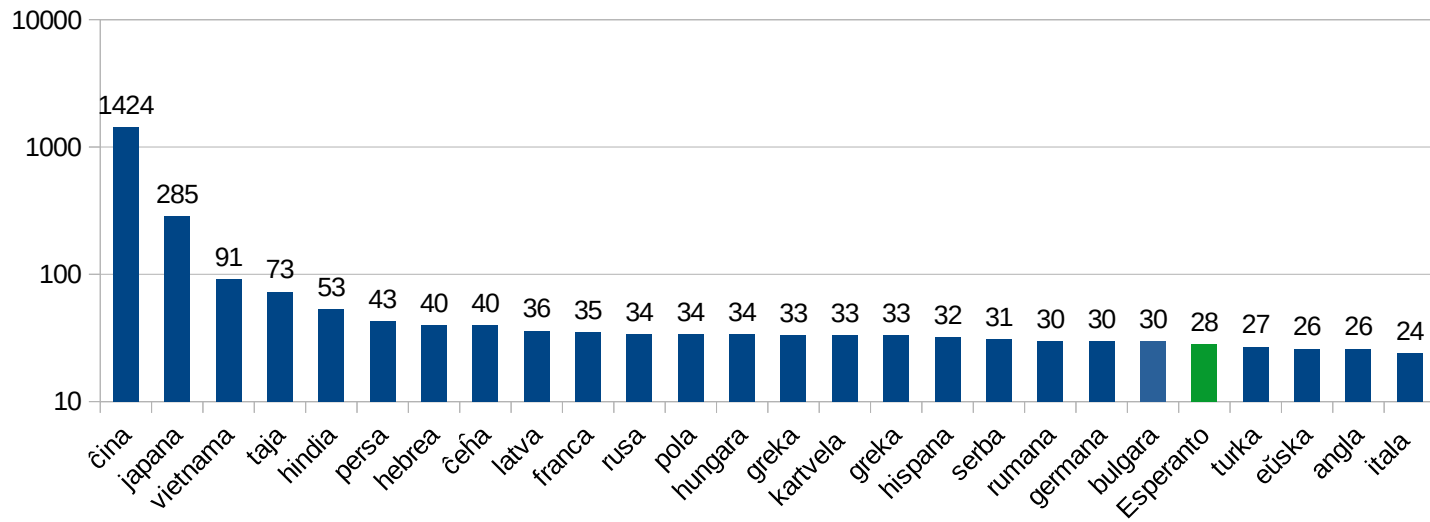
Thai: ที่ฉันเล่าให้คุณฟังถึงรายละเอียดของดาวดวงที่ บี.๖๑๒ และที่ฉันบอกหมายเลขของดาวดวงนี้ก็เพราะพวกผู้ใหญ่คนพวกนี้เขาชอบตัวเลข

Chinese: 我如此不厭其煩地詳細說明遊星 B 六一二號並且告訴你它的號碼，乃是由於大人們的緣故。因為大人們喜歡數目字。

Japanese: こうやって、しょうわくせい B612 のことをいちいちいたり、ばんごうのはなしをしたりするのは、おとなのためなんだ。おとなのひとは、すうじが大すきだ。

In contrast - Esperanto and many other languages separate words with spaces.

The following table shows the number of characters (lowercase letters/ideograms without punctuation marks and without numbers) that you need to know to read this book!



It seems that learning languages written in **Latin alphabets** is advantageous. But one must take into account that these languages use uppercase and lowercase letters and that the written and printed forms of letters are diverse. It can be calculated that one must learn **four times the number of letters**.

On the other hand - Hebrew, Persian, Hindi and Thai letters do not differentiate between lowercase and uppercase letters – written and printed forms of Persian, Hindi and Thai letters are identical.

So an advantage for those languages!

Also interesting is the relationship between written and spoken language.

Esperanto is very **simple** in that sense.

Each letter has one sound – each sound corresponds to only one letter.

The only rule is that the penultimate syllable is stressed.

In other languages it is more complicated.

In French you can easily guess the pronunciation when you see the written word but not vice versa .

In English it is often necessary to learn both the spelling and the pronunciation, because there are no simple rules.

Chinese uses ideograms. Each ideogram corresponds to one syllable.

But the writing has absolutely no relation to speech. When you learn the pronunciation of Chinese syllables you have to learn the four sounds that a syllable has.

Although written in Latin letters, Vietnamese syllables even have six different sounds!

There are many things that can be compared between different languages.

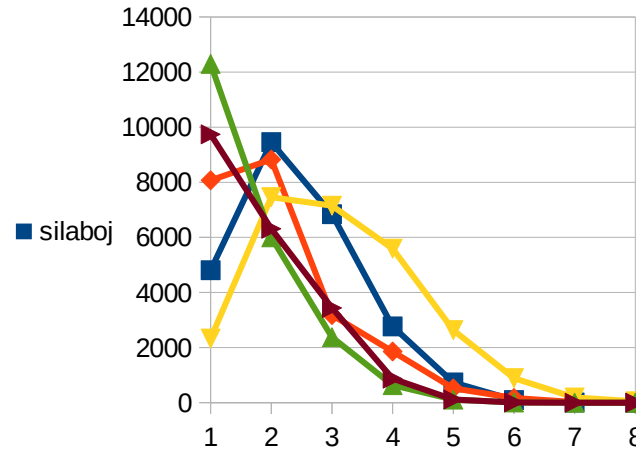
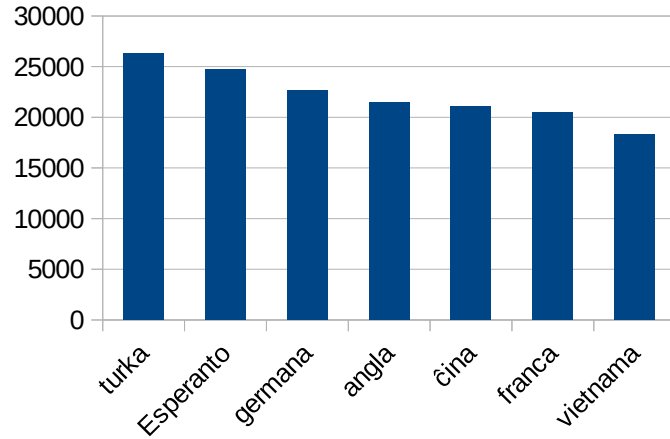
For example:

What are the most common letters used?

How many syllables does the text consist of?

(Thesis: fewer syllables - faster speech
But there are long and short syllables!)

How many letters/syllables does a word have on average?



Esperanto 1,95
Germana 1,60
Turka 2,49
Angla 1,31
Franca 1,42

i a e o n
e n i r s
i a e n r
e t a o i
e i s t a

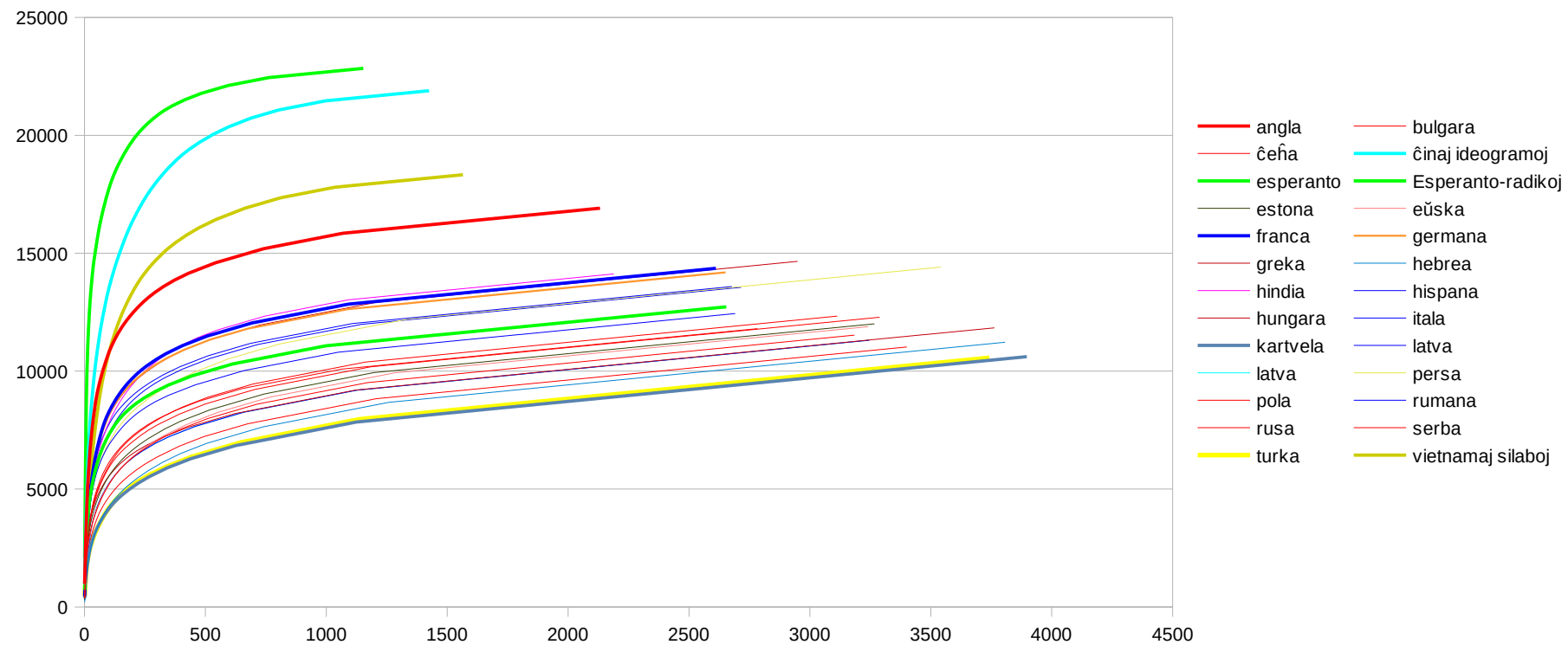
The 5 most common letters

But when learning a foreign language, the number of words you need to learn is very important.

My thesis is that a language is easiest to learn when you have to learn the fewest words!

To investigate this, I counted the words and their frequency and drew a diagram comparing the number of words and the number of different words.

Here is the diagram: The X-axis shows the number of different words and the Y-axis shows the number of words in the entire text of "The Little Prince" .



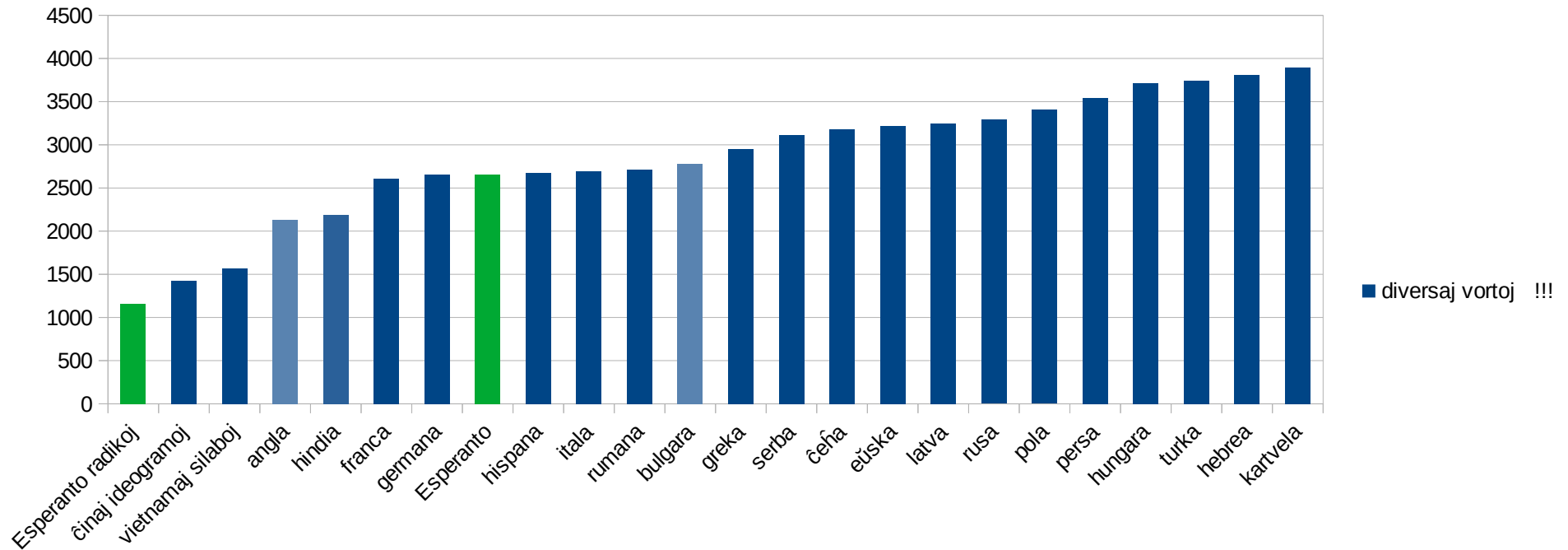
It is observed that languages with more words have fewer different words and conversely, languages with fewer words have more different words.

I try to demonstrate the reason in the table below which contains several translations of the sentence «I am a geographer» .

While Esperanto and French need 3 words, in Czech there are only 2 words because «jsem» already means «I am».

In Turkish you can even express the entire sentence with just one word, namely «coğrafyacığım»!

language	sentence	number of words/ roots	total number	different words
Esperanto roots	Mi est'as geograf'o	5	22842	1154
English	I am a geografer	4	16912	2132
French	Je suis géographe	3	14363	2612
Esperanto	mi estas geografo	3	12732	2652
Czech	jsem zeměpisec	2	11526	3188
Turkish	coğrafyacığım	1	10567	3744



In terms of the number of words to learn, learning Esperanto does not seem to compare favorably with French, German, Spanish, Italian, and Romanian. You have to learn about 2600 words to understand the book "Le petit prince".

It would be advantageous, for example, to learn English words (2133) or to learn Chinese ideograms (1424).

The great advantage of Esperanto is that you can easily divide words into word roots! The diagram shows that the number of Esperanto roots (1154) is the smallest compared to words in other languages!

This can be seen in the following table, which lists the percentage of the text you know when you have learned 500 or 1000 words/roots of a language.

With 500 Esperanto roots you already know 95% of the text, but even with 1000 Turkish words you only know 73% of the text.

	Esperanto roots	Vietnamese		Esperanto		Czech		
		Chinese		English		French		Turkish
500 words/roots/ideograms	95,5%	89,2%	88,0%	85,3%	78,2%	79,6%	68,8%	61,7%
1000 words/roots/ideograms	99,3%	96,7%	96,6%	92,9%	87,0%	88,0%	79,4%	72,6%
100% (number of words)	1154	1456	1569	2132	2652	2616	3188	3748
90%	268	509	571	755	1379	1176	2032	2688
80%	120	269	299	319	573	508	1025	1631

Conclusion:

All translations of "The Little Prince" have the same content - and therefore provide the same information! It is noteworthy that young children all over the world learn to speak and understand their native language very quickly.

Adults find it more difficult to learn foreign languages!

Since I don't know the languages presented here in detail, I don't dare to judge which of them is the easiest language.

Of course, learning a language depends, among other things, on whether you already know something about the foreign language in addition to your native language, such as letters, syllables, words, grammatical structures.

Esperanto is definitely one of the easiest languages!

The two main advantages of Esperanto are:

- a) Each letter corresponds to only one sound. Each sound corresponds to only one letter.
- b) The words are not inflected and agglutinated word roots are easily recognizable.

Here are some internet pages related to the presentation:

<https://www.petit-prince-collection.com>
<https://translate.google.com>

<http://www.petit-prince.at/links.htm>
<https://welt-der-sprache.info>

By the way:

On the previous page you could see the essence of my talk,
that you can only see with your heart ;-)

Thank you for patiently listening to my presentation!

Don't forget!

Never believe statistics that you haven't falsified yourself!

Vertraue keiner Statistik, die Du nicht selbst gefälscht hast!
(popular German slogan)

Contact: uli.ender@esperanto.de