

Entrenament de motors de traducció automàtica estadística entre el castellà i el romanés especialitzats en farmàcia i medicina

Treball de fi de màster

Víctor Peña Irlés, Adrià Martín Mor (director). Màster en Tradumàtica 2017-2018. UAB.

INTRODUCCIÓ I OBJECTIUS

L'objectiu del treball és l'entrenament de motors de traducció automàtica estadística (TAE) entre el castellà i el romanés especialitzats en els camps de la farmàcia i la medicina.

En el cicle complet d'entrenament dels motors s'inclouen diverses motivacions personals i objectius secundaris: cercar recursos en romanés i castellà, aprofundir els coneixements en TAE i enginyeria lingüística, estudiar el comportament dels motors amb el romanés i avaluar la qualitat de la TA en comparació amb altres motors existents.

TRADUCCIÓ AUTOMÀTICA ESTADÍSTICA?

Es basa en models estadístics probabilístics.

Parteix de segments de longitud variable.

Es necessiten corpus:

- Bilingües (model de traducció)
- Monolingües (model de llengua)

CERCA I DESCÀRREGA DE CORPUS

- 1 EMEA (Agència Europea de Medicina)
ECDC (Centre Europeu per a la Prevenció i el Control de les Malalties)
Viquipèdia (per categories)

METODOLOGIA

3

PROCESSAMENT DELS FITXERS

Conversió a format Moses

Macros amb Notepad++

Java

Okapi Tikal

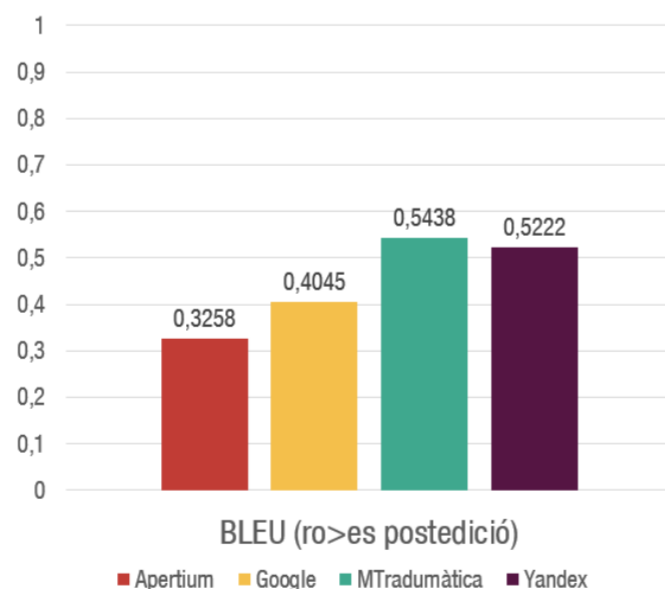
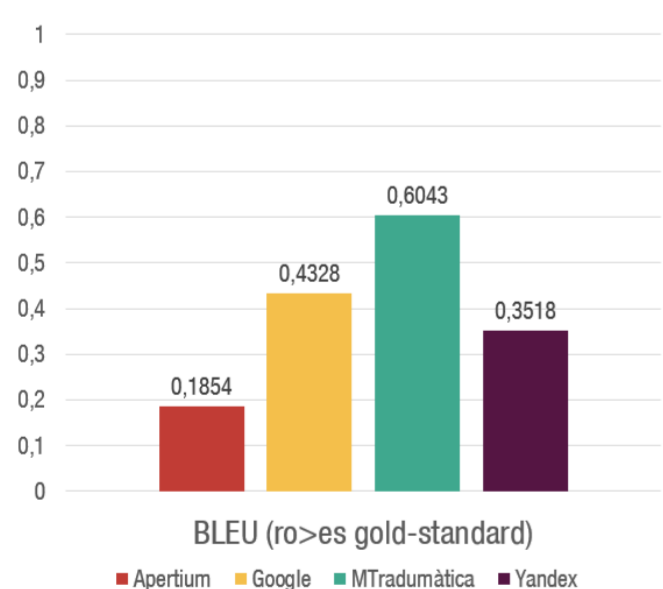
2

ENTRENAMENT A MTRADUMÀTICA

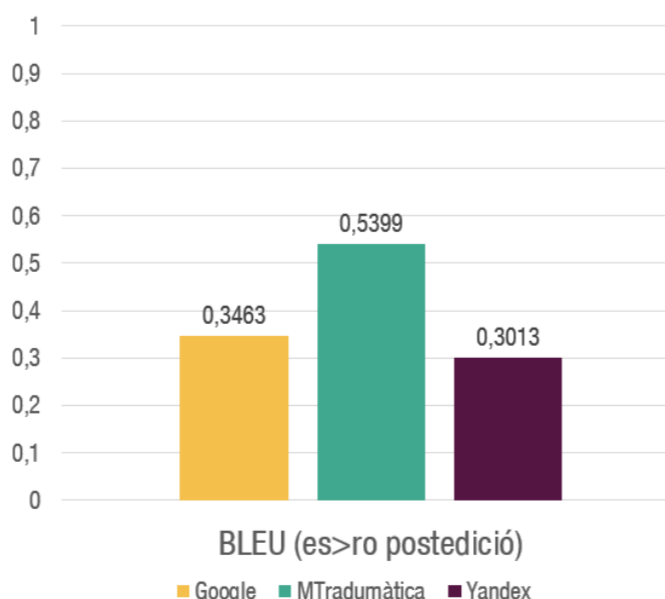
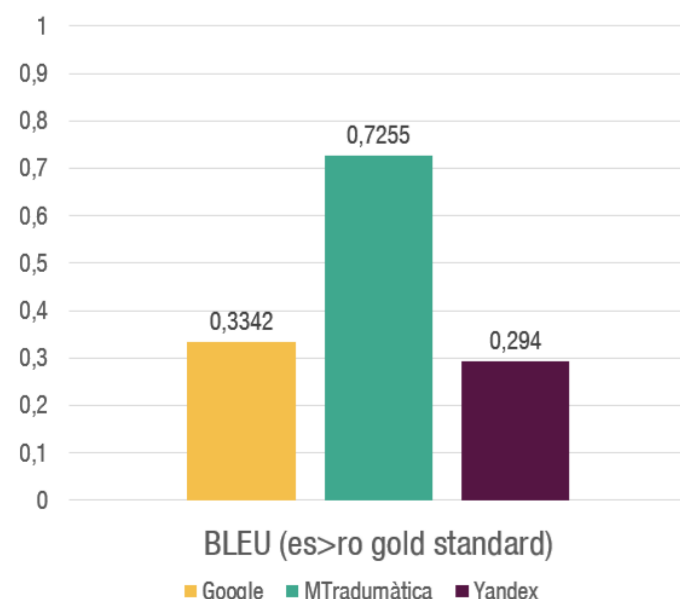


RESULTATS

Romanés-castellà



Castellà-romanés



CONCLUSIONS

Els motors de TAE entrenats registren bons resultats per a textos especialitzats i supera els traductors generals analitzats (Google, Yandex i Apertium)

Es poden entrenar motors de TAE que donen bons resultats amb pocs recursos.

Aquests motors podrien funcionar a la indústria farmacèutica, especialment per a la traducció de prospectes, atès que la posteducació seria totalment viable.

MTradumàtica és una plataforma que acosta la TAE als traductors, atès que funciona d'una manera intuïtiva i pedagògica.

Pel que fa al romanés, cal tenir en compte la codificació dels caràcters diacrítics per a l'entrenament dels motors i la traducció mateixa.