

Euskal post-edizioa, zertan dabil?

Nora Aranberri
IXA Taldea, UPV/EHU

Teknologia itzultzailearen lagun. Noiz eta nola? Jardunaldia.
Eibar, 2016ko ekainaren 22an

Aurkibidea

- Zer da post-edizioa?
 - Teoria eta esperimentuak
 - Trebakuntza eta gidalerroak
- Post-edizio lana errealitate bat al da?
 - Non erabiltzen da? Nola erabiltzen da?
- Zein aldagaiek dute zerikusia post-edizio lanean?
- Euskarazko post-edizioa
 - Euskararako itzulpen automatikoko sistemak
 - Euskararako post-edizio lana
- Ondorioak

Zer da post-edizioa?

Definizioa

Post-edizioa vs. edizioa

Itzulpen automatikoko sistema batek
emandako itzulpen proposamen baten orraztea
aurretik ezarritako kalitate-maila
bermatzeko.

Post-edizioa

Motibazioa:

- Azkarragoa da itzulpen automatiko bat orraztea itzulpen automatikoaren laguntzarik gabe itzultzea baino.
- Ez da itzulpen-kalitatearen galerarik ematen.

Beraz,

IA + PE = itzulpen kopuru handiagoa inbertsio berarekin

Post-edizio lana: adibidea

Itzulgaia

Cargue la batería antes de usarla por primera vez o cuando no se haya utilizado durante mucho tiempo.

IAko proposamena

Bateria kargatu ezazu lehenbiziko aldiz erabili aurretik edo ez da denbora luzez erabili.

Post-edizioa

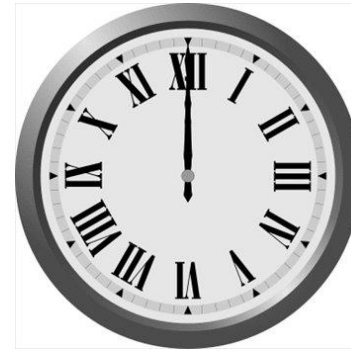
Bateria kargatu ezazu lehenbiziko aldiz erabili aurretik edo ez **bada** denbora luzez erabili.

Zer da post-edizioa?

Teoria eta esperimentuak

Post-edizio lanaren 3 aldeak (Krings, 2001)

1. Esfortzu tenporala



2. Esfortzu kognitiboa



3. Esfortzu teknikoa



Post-edizioa: ikerketa-esperimentuak

1. Produktibitatearen neurketa (Autodesk: Plitt eta Masselot 2010; Gaspari et al, 2014)
2. Esfortzuaren neurketa (Wilker Aziz et al, 2014; Lacruz eta Shreve, 2014; Lacruz et al, 2014)
3. Lan zehatzaren analisisia (Popovic et al, 2013)
4. Kalitatearen neurketa (O'Curran, 2014; Ortiz-Boix eta Matamala, 2015)
5. Post-editorearen profila (Gueberof, 2014; Sánchez-Gijón eta Torres-Hostench, 2014; Aranberri et al, 2014, Mitchell et al, 2013; de Almeida eta O'Brien, 2010)
6. Post-edizio lana vs. itzulpen lana (Elming et al, 2014; Mesa-Lao, 2014)
7. Post-edizio lanaren albo-ondorioak (konsistentzia, itzultzaile berrien lanaren hobekuntza)

Zer da post-edizioa?

Trebakuntza eta gidalerroak

Post-edizioa: Post-editore on baten profila

Bai itzultzaile eta bai post-editoreak hizkuntza baten errepresentatutako esanahia beste hizkuntza batera eramateaz arduratu behar dira.

- Itzultzailearen gaitasunak (Pym, 2003)
 - iturburuko testu batentzat **itzulpen posible** bat baino **gehiago** sortzea;
 - itzulpen horietatik **egokiena** aukeratzea, **azkar** eta **arrazoitzeko** kapaz izanez
- Post-editorearen gaitasunak (Almeida eta O'Brien, 2010)
 - Iaren itzulpenean **zuzendu beharreko** arazoak identifikatu eta behar bezala moldatzeko gaitasuna;
 - post-edizio lana **abiadura dezentez** burutzeko gaitasuna, eguneko produktibitate aurreikuspenak betetzeko modukoa;
 - gidalerroak jarraitzeko gaitasuna, beharrezkoak ez diren ahalik eta **aldaketa gutxien** egiteko, norberaren apetak asetzeko egiten direnak

Post-edizioa: trebakuntza

- Enpresetatik bideratutako trebakuntza
 - Itzulpena automatikoaren historia eta egiturak
 - Erroreen identifikazioa (eta konponketa)

Post-edizioa: trebakuntza



- Unibertsitateko trebakuntza
 - Finlandiatik esperientzia (Koponen, 2015)
 - Alde praktikoa
 - IA sistema desberdinen konparazioa
 - Aurre-edizioa
 - Post-edizioa itzulgaia erabili gabe
 - Kalitate-maila desberdinen produkzioa
 - IAren ebaluazioa

Post-edizioa: trebakuntza



- Unibertsitateko trebakuntza
 - Irlandatik esperientzia (Doherty eta Kenny, 2014)
 - Alde teorikoa
 - Itzulpen memoriak eta tresnak
 - Itzulpen automatikoa
 - Historia, egiturak, ebaluazioa, aurre- eta post-edizioa, hizkuntza kontrolatuak...
 - Alde praktikoa
 - Itzulpen memorien inguruko proiektua
 - IA sistema estatistiko baten garapena eta hobekuntza

Post-edizioa: gidalerroak

- Pribatuak



- Kalitate-mailen araberako errore sailkapena

- Errore sailkapena



**Post-edizio lana errealitate
bat al da?**

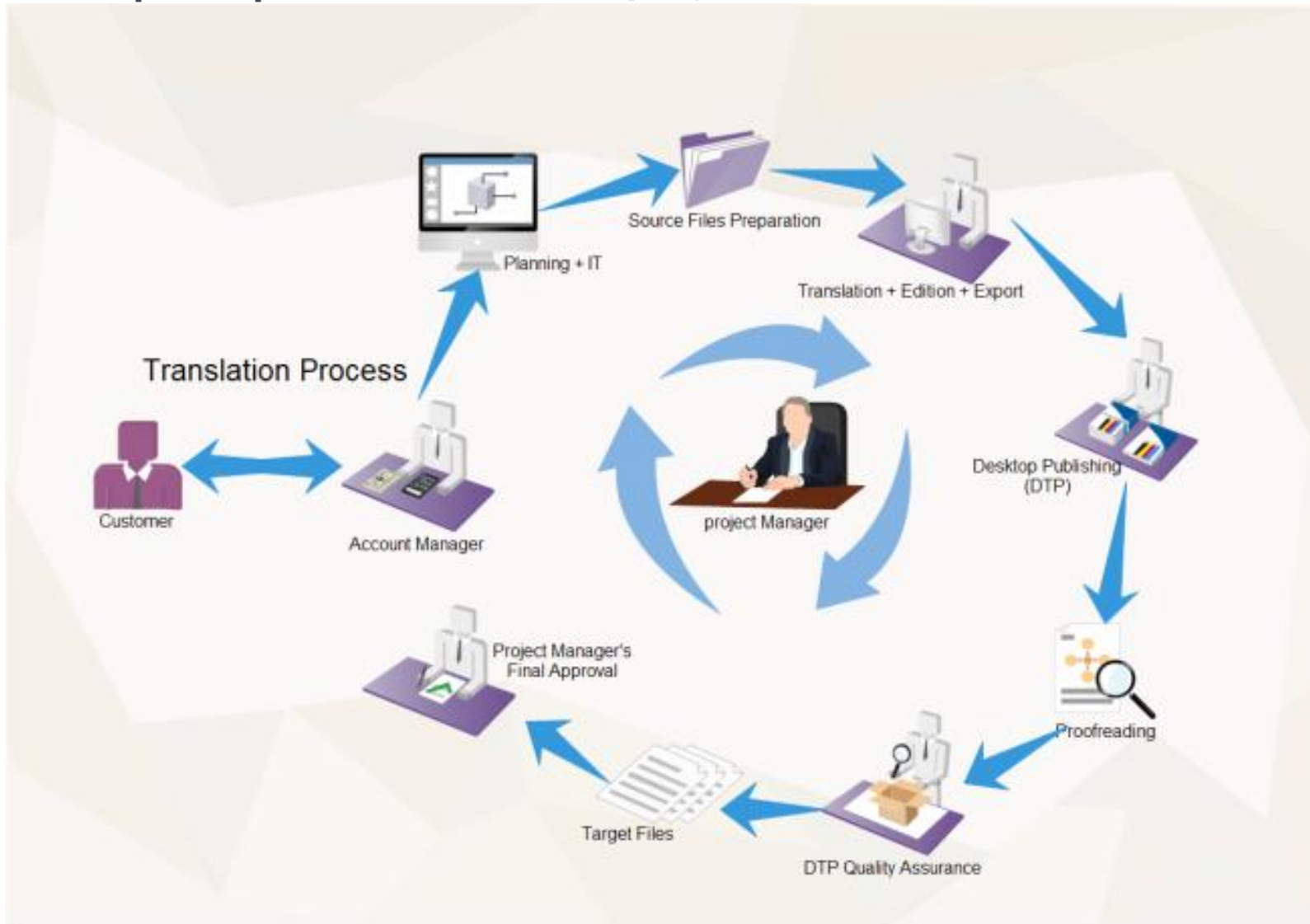
Non eta nola erabiltzen da?

Post-edizioa: non?

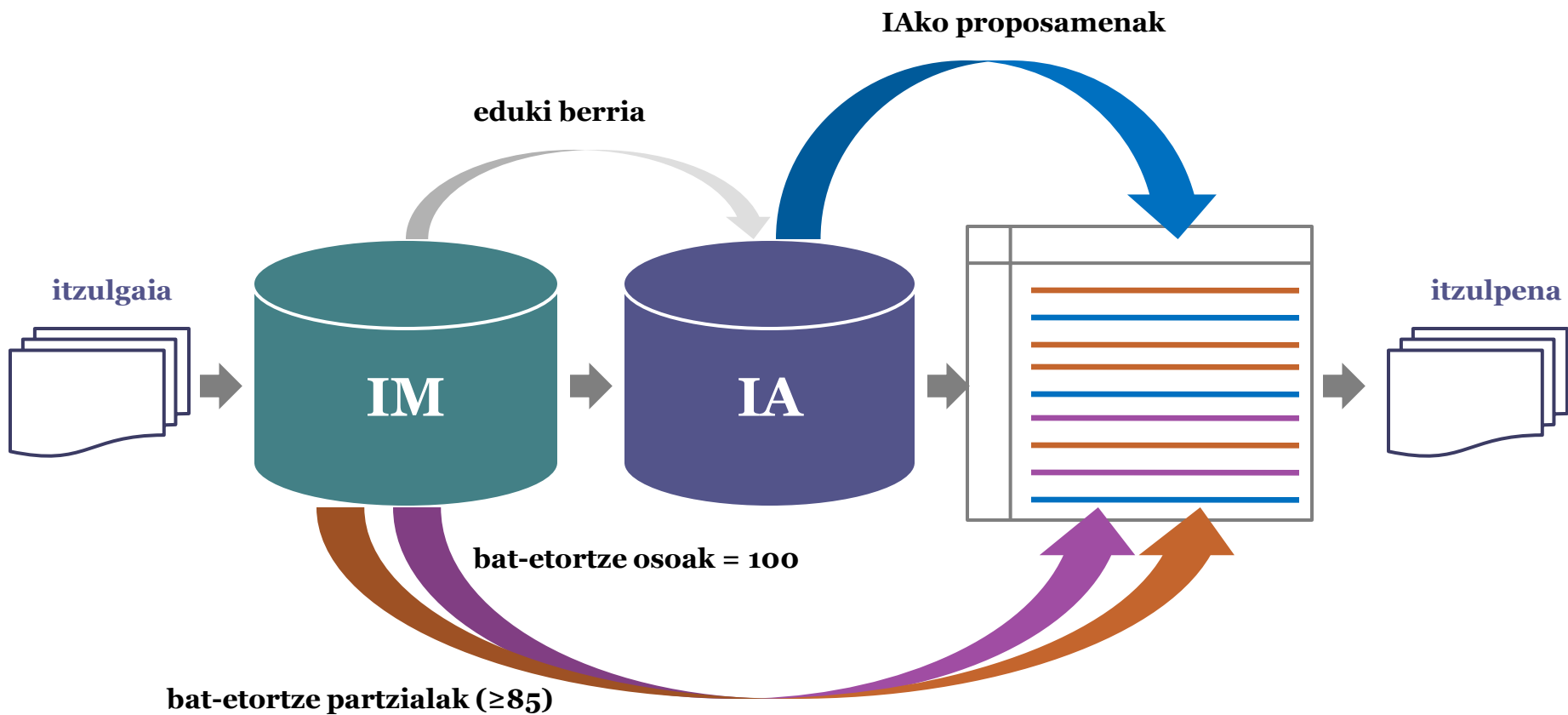


Post-edizioa: nola?

Itzulpen prozesua (ikuspegi orokorra)



Itzulpen prozesua (gertuko ikuspegia)



Itzultzailearen lan-ingurunea: IM + IA

The screenshot displays the Wordfast software interface. The main window shows a translation table with the following data:

ID	French	English	Score
9	12 septembre 2012.	IA itzulpena	IA
10	Appuyez de nouveau sur la touche « HORLOGE » pendant 2 secondes.	Press the "CLOCK" key for 2 seconds again.	84
11	Le dispositif propose une heure de la journée.	The device suggests a time of the day.	100
12	{ut1} Utilisez les touches « Précédent/Suivant » pour sélectionner l'heure souhaitée.	{ut1} Use the "Forward" and "Backward" key to select the desired time.	83
13	Par exemple :	for example:	100
14	9:30.	7:30.	99
15	Vous avez besoin de mettre séparément des heures et des minutes.	You need to put hours and minutes in separately.	100
16	Appuyez sur la touche « HORLOGE ».	IA itzulpena	IA
17	{ut1} Vous pouvez maintenant écouter le réglage de la date et l'heure.	You can now listen to the date and time setting.	99
18	Le réglage de la date et l'heure est complet.	You can now listen to the date and time setting.	60

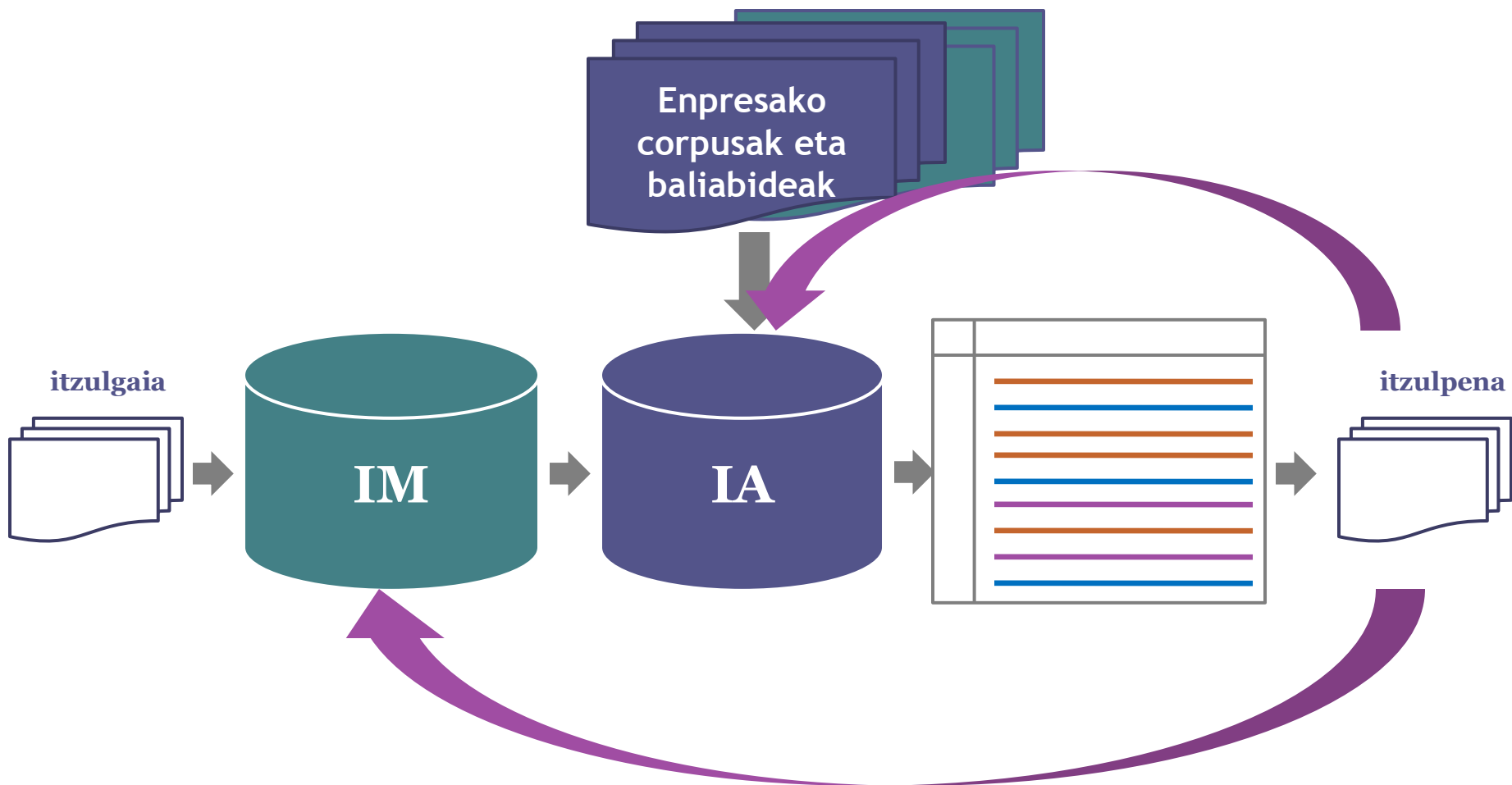
Below the main table, there is a search results table:

ID	Source	Target	Score	TM Name	Date/Time
1	7:30.	7:30.	99	clientTM.txt	8/8/12 3:02 AM

The interface also includes a menu bar (File, Edit, Translation Memory, Terminology, Window, Help), a toolbar with various icons, and an Outline panel on the right side showing a list of items with their corresponding IDs.

Post-edizioa: nola?

- **Pertsonalizazioa** (system customisation)



**Zein aldagaiek dute
zerikusia post-edizio lanean?**

**IA kalitatea, testuaren ezaugarriak,
post-editorearen profila eta gehiago**

Post-edizioa: IAren kalitatea

- Pertsonalizatua vs. Orokorra

Itzulgaia

En **OS X Mail**, elija **Preferencias** > Cuenta, seleccione **la cuenta de iCloud** de la lista de cuentas.

IAko proposamena (sistema orokorra)

OS-etan X. Posta Elektronikoa, **Lehentasunak** aukera ditzan > Kontuak, **kontu-zerrendako iCloud-etako kontua** aukera dezan.

IAko proposamena (sistema pertsonalizatua)

OS X Mail-en, aukeratu **Hobespenak** > Kontua, hautatu **iCloud kontua** kontuen zerrendatik.

Post-edizioa: konplexutasuna

Itzulgaia

No, pero haga una copia de seguridad por si acaso.

IAko proposamena

Ez, baina segurtasun-kopia bat badaezpada egin ezazu.

Itzulgaia

Tengo una cuenta en Youtube de la cual no conozco su correo electrónico asociado o contraseña, me gustaría suprimir videos, y preferiblemente todo el canal. ¿Cómo podría hacer eso?

IAko proposamena

Badut Youtube-ko zeinetako bere posta elektroniko elkartua ezagutzen ez dudan kontua edo pasahitza gustatuko litzateke bideoak, eta ahal izanez gero kanal guztia kentzea. Nola egin lezake hori?

Post-edizioa: Post-editorearen profila

- Itzultzaile-eskarmantua
- Post-edizio-eskarmantua / trebakuntza
- Jakintza-arloarekiko gertutasuna
- IArekiko jarrera

Euskarazko post-edizioa

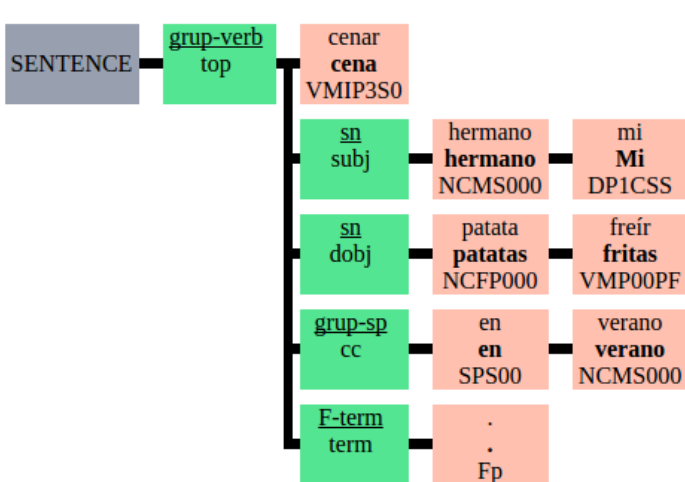
IA sistemak

Euskararako IA sistemak

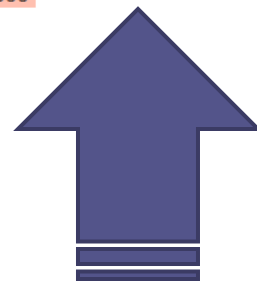
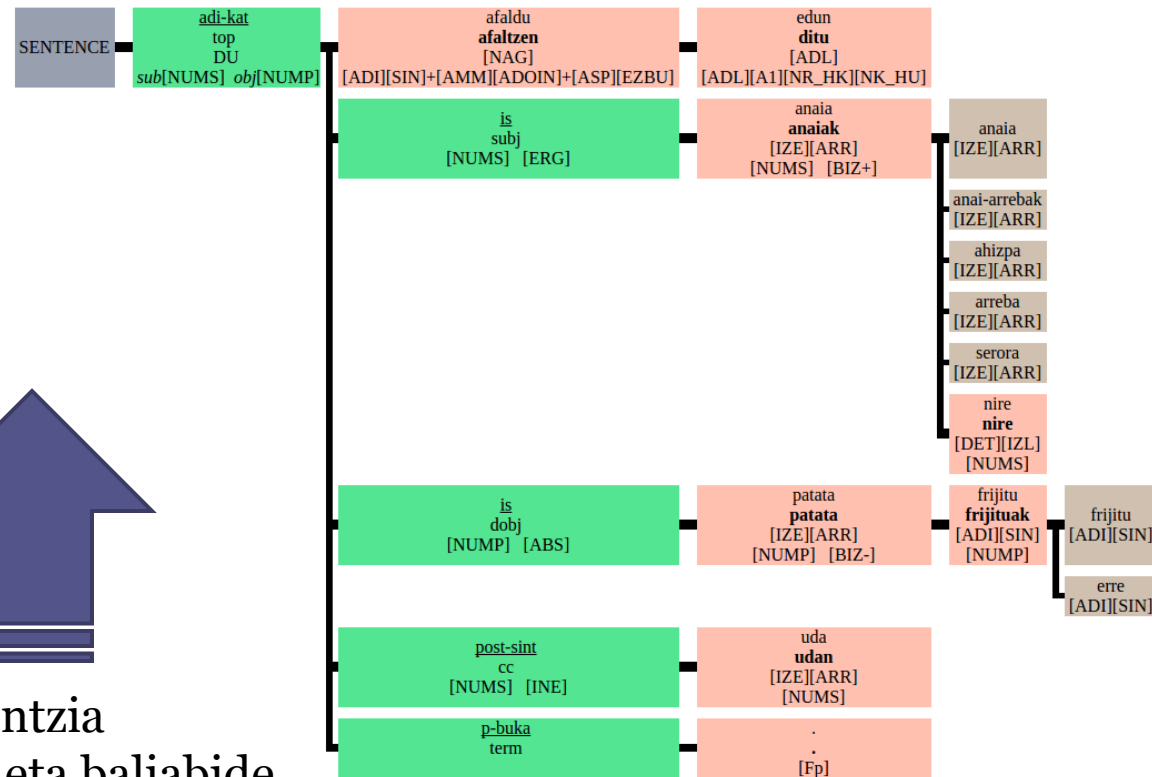


Erregeletan oinarritutako sistemak

Mi hermano cena patatas fritas en verano .



nire anaiak patata frijituak afaltzen ditu udan .



Transferentzia
erregelak eta baliabide
linguistikoak

analisi-zuhaitza

sorkuntza-zuhaitza

Estatistikan oinarritutako sistemak

Probabilitateen itzulpen taula

joins several text strings ||| hainbat test-kate elkartzen ditu ||| 6.41133e-05
joins several text strings ||| zenbait test-kate elkartzen ditu ||| 5.3741e05
joins several text ||| hainbat ||| 1.03783e-09
joins ||| ! ||| 1.90793e-05
repeats text ||| testua errepikatzen du , adierazi ||| 0.0442171
repeats text ||| testua errepikatzen du , ||| 0.0442171
repeats text ||| testua errepikatzen du ||| 0.0442171
replaces characters ||| karaktereak ordeztzen ditu ||| 0.0835947
replaces part of a text ||| Testu-katearen zati bat ||| 4.04003e-06
replaces part of a ||| Testu-katearen zati bat ||| 0.0118755
replaces part of ||| Testu-katearen zati ||| 0.0206278
replaces ||| Testu-katearen ||| 0.317989

Hizkuntza eredua

-0.5295186 sakoneko ura gainazalera
-2.4733522 , gorputzaren gainazalera
-1.2987162 <s> izpiak gainazalera
-0.82516205 zen lurraren gainazalera
-0.52691126 gainazaletik lurraren gainazalera
-0.52501 edo Eguzkiaren gainazalera
-0.87118936 ur sakonak gainazalera
-1.5456117 eta Lurraren gainazalera
-1.1412746 du Lurraren gainazalera
-1.680263 , Lurraren gainazalera
-1.0604944 bat Lurraren gainazalera
-0.51076365 berriro Lurraren gainazalera
-0.5266557 bat aintziraren gainazalera
-0.52855456 urtzaldiek lurzorua gainazalera
-2.2474844 eta zelularen gainazalera
-1.0013907 konplexua zelularen gainazalera
-0.5266976 horretatik larruazalaren gainazalera
-0.528472 atmosferako karbonoa gainazalera
-1.3748051 da ohartzea gainazalera
-0.5268555 azalpeko urari gainazalera
-0.9856331 lurraren atmosferak gainazalera
-0.5104105 mahaziak putzuaren gainazalera
-0.5189849 ur hotzagoa gainazalera
-0.43285996 gutxiko zorupeak gainazalera
-0.16159998 kaltzita-erronboaren azpialdeko gaina
-1.4082879 , epitelioaren gainazalera
-0.35372564 zelula antigenoaurkezleen gainazalera
-0.823043 mineral aleen gainazalera

Euskarazko post-edizioa

Post-edizio lana

UEU post-edizio lantegia (2015eko udazkena)

- **Helburua:** euskarazko post-edizioaren lehen hurbilpena egitea, post-edizio lana eta eztabaida uztartzeko foroa sortzea
- **Testuingurua:**
 - 10 itzultzaile profesional
 - 8 aste
 - EJko itzultzailea (bereziki) + Emakundeko txostena (bereziki)
 - 5 saio presentzial (alde teorikoak jorratzeko eta aurreko lanari buruz hausnartzeko)
 - asteko 4 post-edizio ataza eta produktibitate ataza 1
 - zenbait segmenturen eskuzko ebaluazioa

UEU post-edizio lantegia: ondorioak

- **Euskararako IA sistema orokorren kalitate kaskarra eta aldakorra**
 - **IA sistemen pertsonalizazioaren beharra**
- **Post-editoreak sisteman eragiteko gaitasunaren komenigarritasuna**
 - **Lexikoa gehitzeko aukera**
 - **Egitura zehatzen itzulpena zuzentzeko aukera**
- **Kalitate kaskarreko proposamenekin lan egiterakoan, post-edizio lana ez da errorean orrazketa bezala tratatzen**
 - **Atazaren deskribapena: proposameneko azpi-segmentu egokien identifikazioa, norberaren ekarpenen gehitzea eta elementu guztiek “jostea”**
- **Kalitate kaskarreko proposamenekin lan egiterakoan, post-edizio lana ez da errorean orrazketa bezala tratatzen**
 - **Nahitaezkoak ez diren aldaketak ekartzen dira**
- **Kalitate “onargarria” zehazteko zailtasuna**
- **IA pizgarri bezala**
 - **Post-edizioak eta IA gabeko itzulpenak orokorrean denbora bera darama**
 - **8 astetako lanaren ondoren ez da aldaketarik nabari, esaldi luzeetan ezik**
 - **Ez da kalitatean galerarik nabaritzen orokorrean**

Ondorioak

Ondorioak

- Post-edizioa, itzulpen automatikoko sistemen proposamenen orrazketa, gero eta hedatuago dagoen ataza da.
- Post-edizioaren inguruan ikerketa ugari burutzen hasi da azken urteotan ataza ulertzeko asmoz.
- Post-edizioan trebatzea beharrezkoa dela onartu da baina zehaztugabe dago oraindik nola jorratu beharko litzatekeen.
- Euskararako post-edizioa abian dago eta sistemen pertsonalizazioaren bidez lortuko da bultzada ematea.
 - Euskal post-editoreak trebatu beharra izango dugu hurrengo urteetan.

Bibliografia

- Hans P. Krings. 2001. Repairing texts: Empirical investigations of machine translation post-editing process. The Kent State University Press, Kent, OH.
- Plitt, M. eta Masselot, F. (2010): “A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context”, in The Prague Bulletin of Mathematical Linguistics, 93, 7-16.
- Zhechev, V. (2012): “Machine Translation Infrastructure and Post-editing Performance at Autodesk” in Proceedings of the AMTA 2012 Workshop on Post-editing Technology and Practice (WPTP’12).
- de Almeida, G. eta O’Brien, S. (2010): “Analysing post-editing performance: correlations with years of translation experience”, in Proceedings of the 14th annual conference of the European association for machine translation, St. Raphaël, France.
- DePalma, D., Hegde, V., Pielmeier, H. eta Stewart, R. (2013): The Language Services Market: 2013. Common Sense Advisory, Boston, USA.
- DePalma, D., Pielmeier, H., Henderson, S. eta Stewart, R. (2015): The Language Services Market: 2015. Common Sense Advisory, Boston, USA.
- Cronin, M. (2013): Translation in the Digital Age. Routledge, Oxfordshire, UK.
- Guerberof, A. (2014): “The Role of Professional Experience in Post-editing from a Quality and Productivity Perspective”, in S. O’Brien, L. Winther Balling, M. Carl, M. Simard eta L. Specia (arg.), Post-editing of Machine Translation: Processes and Applications, Cambridge Scholars Publishing, Newcastle upon Tyne.
- Elming, J., Winther Balling, L. and Carl, M. (2014): “Investigating User Behaviour in Post-editing and Translation using the CASMACAT Workbench”, in S. O’Brien, L. Winther Balling, M. Carl, M. Simard eta L. Specia (arg.), Post-editing of Machine Translation: Processes and Applications, Cambridge Scholars Publishing, Newcastle upon Tyne.
- Aziz, W., Koponen, M. and Specia, L. (2014): “Sub-sentence Level Analysis of Machine Translation Post-editing Effort”, in S. O’Brien, L. Winther Balling, M. Carl, M. Simard eta L. Specia (arg.), Post-editing of Machine Translation: Processes and Applications, Cambridge Scholars Publishing, Newcastle upon Tyne.

- Mesa-Lao. (2014): “Gaze Behaviour on Source Texts: An Exploratory Study comparing Translation and Post-editing”, in S. O’Brien, L. Winther Balling, M. Carl, M. Simard eta L. Specia (arg.), Post-editing of Machine Translation: Processes and Applications, Cambridge Scholars Publishing, Newcastle upon Tyne.
- Lacruz, I. and Shreve, G. (2014): “Pauses and Cognitive Effort In Post-editing”, in S. O’Brien, L. Winther Balling, M. Carl, M. Simard eta L. Specia (arg.), Post-editing of Machine Translation: Processes and Applications, Cambridge Scholars Publishing, Newcastle upon Tyne.
- Koponen, M. (2013): This translation is not too bad: An analysis of post-editor choices in a machine translation post-editing task. Workshop on Post-editing Theory and Practice 2013.
- Popovic, M., Avramidis, E., Burchardt, A., Vilar, D. and Uszkoreit, H. (2013): What can we learn about selection mechanism for post-editing? Workshop on Post-editing Theory and Practice 2013.
- Mitchell, L., Roturier, J., and O'Brien, S. (2013): Community-based post-editing of machine-translated content: monolingual vs. bilingual. Workshop on Post-editing Theory and Practice 2013.
- Aranberri, N., Labaka, G., Diaz de Ilarraza, A., and Sarasola, K. (2014): Comparison of post-editing productivity between professional translators and lay users. Workshop on Post-editing Theory and Practice 2014.
- Gaspari, F., Toral, A., Naskar, S., Groves, D. and Way, A. (2014): Perception vs Reality: Measuring Machine Translation Post-Editing Productivity. Workshop on Post-editing Theory and Practice 2014.
- Lacruz, I., Denkowski, M. and Lavie, A. (2014): Cognitive Demand and Cognitive Effort in Post-Editing. Workshop on Post-editing Theory and Practice 2014.
- O'Curran, E., (2014): Translation Quality in Post-Edited versus Human-Translated Segments: A Case Study . Workshop on Post-editing Theory and Practice 2014.
- Sánchez-Gijón, P. and Torres-Hostench, O. (2014): MT Post-editing into the mother tongue or into a foreign language? Spanish-to-English MT translation output post-edited by translation trainees. Workshop on Post-editing Theory and Practice 2014.
- Ortiz-Boix, C. and Matamala, A. (2015): Quality Assessment of Post-Edited versus Translated Wildlife Documentary Films: a Three-Level Approach. Workshop on Post-editing Theory and Practice 2015.
- Koponen, M. (2015): How to teach machine translation post-editing? Experiences from a post-editing course. Workshop on Post-editing Theory and Practice 2015.
- Doherty, S. and Kenny, D. (2014). The design and evaluation of a Statistical Machine Translation Syllabus for translation students. *The Interpreter and Translator Trainer*, 8(2):276–294.

Eskerrik asko!

nora.aranberri@ehu.eus