

Itzulpen automatikoa wikilariak laguntzeko

Iñaki Alegria eta Kepa Sarasola

Wikipedia-Wikimedia editatoia 2020

[Euskal Wikipediako wikiproiektua](#)

[UEUko udako ikastaroa.](#)

Aurkibidea

Itzulpen automatikoa wikilariak laguntzeko.....	1
SARRERA.....	2
Itzulpen Automatiko Neuronala.....	2
Euskarazko testuetarako Itzulpen Automatikoa.....	2
NOLA ERABILI CONTENT TRANSLATION TRESNA?.....	6
1. aldia.....	6
Gainontzekoetan.....	6
Artikulu bat itzultzeko urratsak. Estrategia.....	6
ADIBIDE BAT.....	7
KONTUZ IRUDIEKIN.....	10
KONTUZ ERREFERENTZIEKIN.....	10

SARRERA

Itzulpen Automatiko Neuronal

Itzulpen automatikoa (IA edo *MT-Machine Translation*) testuen itzulpena automatikoki lortzeko ahalbidetzen duen teknologia da. Teknologia honek 1940tik hainbat hamarkadaz laguntza mugatua eskaini du, baina 2017tik arrakasta handia lortu du, eta harrezkero erabilera askoz zabalagoa da.

Dena dela itzulpen automatikoaren emaitzak ez dira inoiz perfektuak izan, eta oraindik ere ez dira perfektuak. Emaitza horiek publikoki zabaldu nahi badira, guztiz beharrekoa da pertsona batek testua zuzentzea eta orraztea ([post-edizioa](#)).

[Itzulpen automatiko neuronalaren aurrerapenak eskura](#) artikuluan esaten den bezala: *sare neuronalen eta ikasketa sakonaren bitartez itzulpen automatikoan aurrerapen handiak gertatzen ari direla. Hobekuntza itzulpenaren naturaltasunean nabarmentzen da, batez ere, irakurtzeko eta ulertzeko errazagoak diren testuak lortzen baitira.*

Itzulpen automatikoari aplikaturiko metodoak historikoki hiru tetapatan bana ditzakegu:

- [Erregeletan oinarritutako IA](#) (1940an asmatua). Era horretakoa da 2006an kaleratu zen Matxin itzultzailea.
- [Corpusetan oinarrituriko IA](#) (2000an asmatua). Era horretakoa da EUSMT eta Euskoa Jaurlaritzak orain dela gutxi arte erabiltzen zuen itzultzaile gramatikala
- [Itzulpen automatiko neuronal](#) (2014an asmatua)

Erregeletan eta corpusetan oinarritutako itzultzaileek laguntza ederra eskaini dute hamarkadetan pertsoneri, baina beti ere oraingo kalitate mailatik urruti. 2014z geroztik [neurona-sare](#) handiak erabiltzen hasi ziren hitz-sekuentzia batek duen agertzeko probabilitatea aurreikusteko, eta normalean esaldi osoak ere modelatzen dira eredu integratu bakar batean. Itzulpen automatiko neuronal sakona aurrekoaren hedadura bat da. Biek erabiltzen dute neurona-sare handi bat baina itzulpen automatiko neuronal sakonak hainbat geruza prozesatzen ditu batera bakarka egin ordez. Alor honetan izan den garapena ikaragarri azkarra izan da, lorpenak eta hobekuntzak arin-arin gertatu dira urte gutxi-gutxi batzuetan. Horrela, [Ikaskuntza sakona](#) (deep learning) teknika Itzulpengintza automatikoan erabili zuen lehenengo argitalpen zientifikoa [2014ean](#) agertu zen, egileak [Bahdanau](#), Cho eta Bengio izan ziren. Handik urtebetera [2015ean](#) bazegoen neurona-sistema bat itzulpen automatikozko txapelketa batean (OpenMT'15). Eta hurrengo urtean txapeldunen %90a sistema neuronalak izan ziren. 2015eko aurrerapena arretan oinarritutako sistemak ("attention based NMT") erabiltzea izan zen. 2017an hobekuntza gehigarriak lortu ziren [neurona-sareetan Transformer arkitektura](#) erabiltzearen ondorioz. Urte bat geroago euskara ere arrakastaz erabiltzen zuen sistema bat zegoen, eta beste urte batean, 2019an, euskararako zeuden neurona-sistemak bost ziren.

Euskarazko testuetarako Itzulpen Automatikoa

Paradigma neuronalaren ekarriarekin 2017z geroztik izugarrizko hobekuntza nabaritu zen hizkuntza nagusien arteko itzultzaile automatikoetan. Geroxeago, eta arin, euskal munduko ikerkuntza komunitatea gai izan zen euskaratik eta euskararako itzultzaile neuronalak mundu mailako artearen egoeraren pare jartzeko.

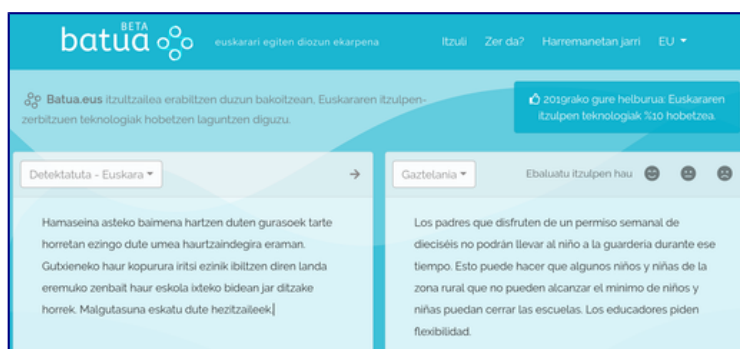
2015ean hasi zen euskararako itzulpen neuronalaz aztertzen TADEEP proiektuan (<http://ixa2.si.ehu.es/tadeep/>) Ordurako [DeepL](#) itzultzaileak kalitatezko emaitzak ematen zituen 10 hizkuntzatan baina euskara ez zegoen horien artean. Bi urte geroago lehenengo emaitza bikainak lortuta, 2017an lehen demoa publikoki eskura zegoen. 2017an bertan, hainbat agente ([Ixa Taldea](#), [Elhuyar](#), [Vicomtech](#), Ametzagaña, [Mondragon Lingua...](#)) batera, MODELA izeneko proiektua abiatu zen (<http://modela.ametza.com>). Beste urtebeteko epean, 2018an MODELA itzultzailea plazaratu zen, Interneten publiko orokorrerako euskarazko itzulpen neuronalaz eskaintzen zuen lehen zerbitzua.

Abiada bizian mugitzen zen arlo honetan, geroago gutxienez beste hiru itzultzaile neuronal gehiago plazaratu ziren. Horien arteko diferentzia nagusia itzultzaile bakoitzak erabiltzen duen corpusa da (corpus = milioika esaldi bi hizkuntzatan), baita itzultzeko zerbitzuan onartzen duten hitz-kopuru maximoa ere (konputagailu ahaltzu eta garestiak erabili behar dira horretan). Hauek ziren 2020ko hasieran eskaintzen ziren euskararako itzultzaile neuronalak:

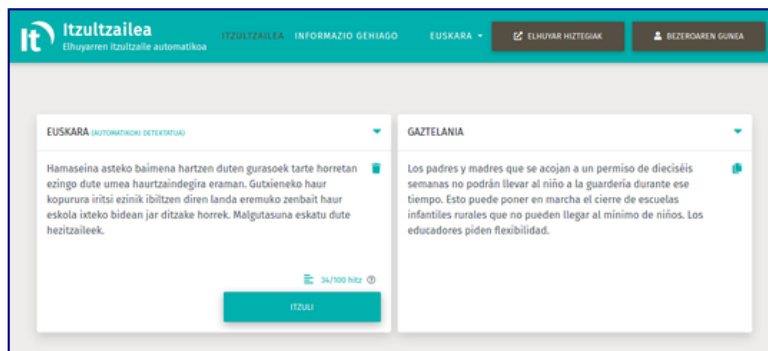
- Eusko Jaurlaritzaren [Itzultzaile neuronalak](#): itzulpen neuronalaren teknologia erabilita, corpus-oinarria Eusko Jaurlaritzako itzulpen-memoriak erabili ziren IVAP-HAEEK 20 urtetan bildutakoa 10 milioi "esaldi" baino gehiago.



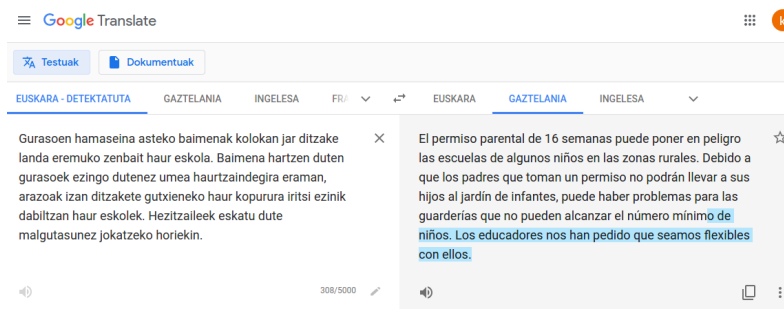
- [batua.eus](#): Vicomtech-ek MODELA sisteman hobekuntzak egin zituen ([RNN](#) teknologiatik [Transformer](#) teknologiarara pasa ziren) eta corpusa handitu zuten.



- [Itzultzailea.eus](#): Elhuyarrek ere antzeko hobekuntzak egin zituen bere aldetik. Eta itzulpenak egiteko hizkuntza berriak gehitu zituen (ingeleza, frantsesa, espainiera, galegoa, eta katalana)



- [Translate Google](#): Googleren Translate zerbitzuak hobekuntzak lortu zituen eredu neuronalaren erabilera.



WIKIPEDIAN ARTIKULUAK SORTZEKO LAGUNTZA: *CONTENT TRANSLATION (CT)*

Itzulpen automatikoaren emaitza zuzenean erabil daiteke helburua ulermen hutsa denean, baina, kontuz, helburua testua argitaratzea denean, lortutako testua moldatu behar da zuzentasuna ziurtatzeko eta egokitasuna areagotzeko. Prozesu horri [post-edizioa](#) deitzen zaio. WPren kasuan post-edizio prozesua asko errazten da [Content Translation](#) tresnari esker. Horren bitartez, testuaren itzulpen automatikoa eta post-edizioa bideratzeaz gain, [metadatu](#)en (irudiak, kanpo estekak, wikiestekak, formatuak, erreferentziak...) lokalizazioa ere aurkezten zaigu informazio hori zuzentzeko edo osatzeko aukera eskainiz.

ContentTranslation izeneko funtzionalitateari esker eroso eta eraginkorra da itzulpen automatikoa erabiltzea artikulua berriak gehitzeko Euskal Wikipedian.

Zoritxarrez Wikipedian euskararekin lan egiteko integratuta dagoen itzultzaile neuronal bakarra, oraingoz, Google-koa da, eta momentuz hori erabiliko dugu. Eta ez da hori hoberen dabilena, bertakoak hobexeko ibiltzen dira. Ea laster eskura izango ditugun!

Gaiei dagokienez, nazioarteko biografiak/liburuak/filmak... zein gai tekniko-zientifikoak dira itzulpen automatikoaz itzultzeko egokienak, arlo horietan hizkuntza eta kulturen artean alde txikiagoak daudelako. Edozein kasutan, jatorrizko artikulua gustukoa baduzu, itzulpena egin dezakezu. Geroago Wikilarien komunitateak ikusiko du bere horretan utzi, edo zer osatu, moldatu...

Hasteko euskaraz landu gabe dauden artikulua dira hauta daitezkeenak. Ez dituzu paragrafo guzti guztiak itzuli behar. Nahikoa dela uste duzuenean, argitaratu eta joan Wikipediara bertako editorearekin beste orrazketa ematera.

ContentTranslation itzultzailearekin zaudela lana guztiz amaitu gabe utz dezakezu lan-saioa.. Gero berrartzeko, edo deskonexioren bat gertatu bazaizu, funtzio honen gunera joan zaitezke: <https://eu.wikipedia.org/wiki/Berezi:ContentTranslation>

Irudiek, taulek, estekek, bibliografiak eta txantilo baten arabera dabilzan beste elementu batzuek arazoak eman ohi dute (ik. geroago), beraz, orrazketan horiek ondo aztertu eta, beharrezkoa denean, aldaketak egin.

Itzulpen-sistema honen erabileraren inguruko estatistikak kontsulta daitezke: <https://eu.wikipedia.org/wiki/Berezi:ContentTranslationStats>

Orain arte, mundu osoan guztira 680.619 artikulua itzuli dira, eta horietatik 3.198 euskaraz. 2020ko ekainaren hirugarren astean 4.149 artikulua itzuli dira mundu mailan, eta horietatik 13 itzuli dira euskarara.

ADIBIDE BAT

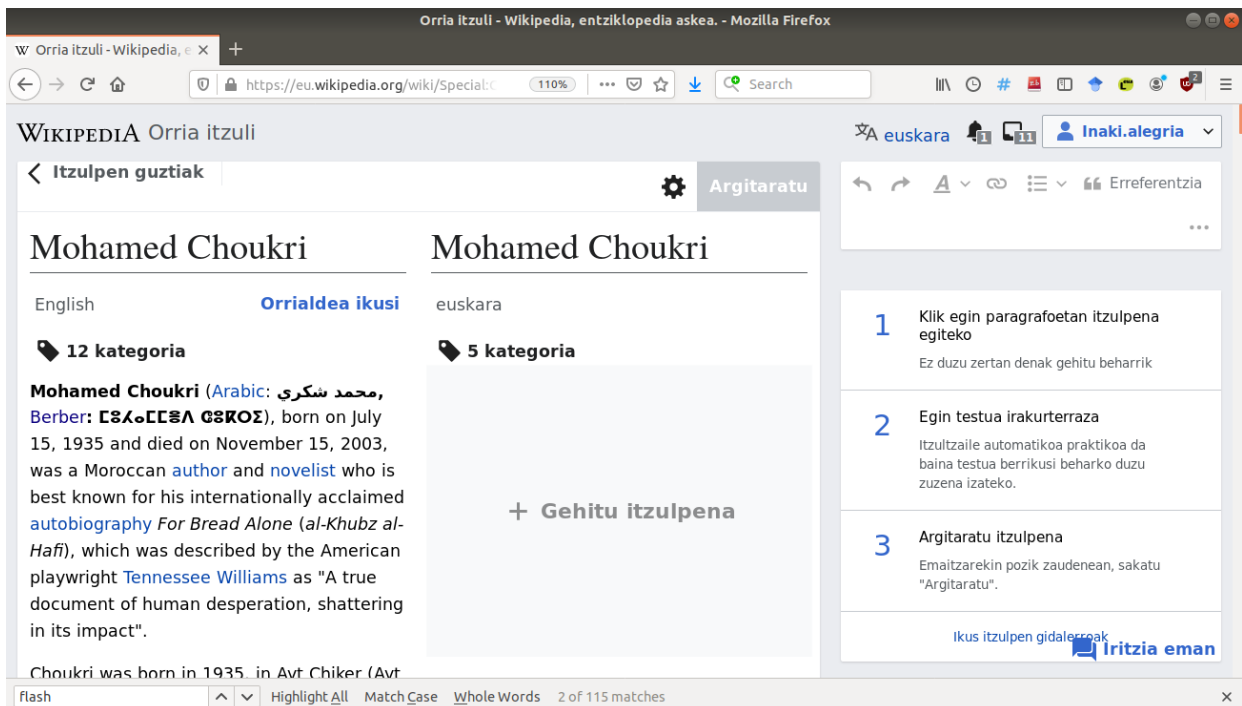
Mohamed Xukriren *Ogi Hutsa* liburua ari naiz irakurtzen eta idazlea ez dago Euskal WPn.

1. Bilatu itzuli nahi duzun artikulua: https://en.wikipedia.org/wiki/Mohamed_Choukri
2. Ondo aztertu artikulua: Indizea garrantzitsua da. Hasiera batean "quotations" kanpoan utziko dut eta filmak agian ere bai.

Contents [hide]

- 1 [Early years](#)
- 2 [Learning how to read and write](#)
- 3 [Censorship of *For Bread Alone*](#)
- 4 [Later life](#)
- 5 [Films](#)
- 6 [Quotations](#)
- 7 [Works](#)
- 8 [See also](#)
- 9 [References](#)

3. Sartu zaitetz itzultzera: euskara-n klik eginez



The screenshot shows the Wikipedia 'Orria itzuli' (Translate) interface. The browser window title is 'Orria itzuli - Wikipedia, entziklopedia askea. - Mozilla Firefox'. The address bar shows the URL 'https://eu.wikipedia.org/wiki/Special:'. The page is in Euskara. The main content area shows the article for 'Mohamed Choukri' in English on the left and Euskara on the right. The Euskara version is currently blank, with a '+ Gehitu itzulpena' (Add translation) button. A sidebar on the right contains instructions for using the tool:

1. Klik egin paragrafoetan itzulpena egiteko
Ez duzu zertan denak gehitu beharrik
2. Egin testua irakurteraza
Itzultzaile automatikoa praktikoa da baina testua berrikusi beharko duzu zuzena izateko.
3. Argitaratu itzulpena
Emaltzarekin pozik zaudenean, sakatu "Argitaratu".

At the bottom of the sidebar, there is a link 'Ikus itzulpen gidalekoak' and a button 'Iritzia eman'.

4. Paragrafoz paragrafo joan. Erabaki paragrafoa txertatu ala ez. Behin txertatuta, osorik ez bazaizu interesatzen, kendu hortik kendu beharrekoa.
5. Zuzendu itzulpena, jatorrizkoa kontuan hartuz.

Orria itzuli - Wikipedia, entziklopedia askea. - Mozilla Firefox

WIKIPEDIA Orria itzuli

Itzulpen guztiak Oraintxe bertan gordea Argitaratu

Mohamed Choukri

English **Orrialdea ikusi** euskara

12 kategoria

Mohamed Choukri (Arabic: محمد شكري, Berber: ⵎⵓⵎⵎⴰⵎ ⵉⵔⵓⵎⵓⵔⵉ), born on July 15, 1935 and died on November 15, 2003, was a Moroccan author and novelist who is best known for his internationally acclaimed autobiography *For Bread Alone (al-Khubz al-Hafi)*, which was described by the American playwright Tennessee Williams as "A true document of human desperation, shattering in its impact".

Choukri was born in 1935. in Avt Chiker (Avt

Mohamed Xukri

euskara

5 kategoria

Mohamed Choukri (arabieraz: محمد شكري, amazigeraz: ⵎⵓⵎⵎⴰⵎ ⵉⵔⵓⵎⵓⵔⵉ), Rif aldeko idazle eta nobelagilea izan zen, 1935ko uztailaren 15ean jaioa eta 2003ko azaroaren 15ean hila. Nazioartean ezaguna da bere autobiografia *den Ogi Hutsa (al-Khubz For al-Hafi)* liburu dela-eta, Tennessee Williams dramaturgo estatubatuarak "gizakiaren etsipenaren egiazko dokumentu" gisa deskribatu zuena.

flash Highlight All Match Case Whole Words 2 of 115 matches


- Zaindu elementu bereziak: irudien oinak, estekak, taulak... Zaila jartzen bada, ez eman denbora gehiegi, bukaeran Wikipedian bertan egin beharreko orrazketan berriro saia zaitezke.
- Esan bezala, luzatzen bazara, utz dezakezu eta berriro hartu URL honetan: <https://eu.wikipedia.org/wiki/Berezi:ContentTranslation>

WIKIPEDIA Itzulpenak

+ Itzulpen berria

Itzulpena egiten

Mohamed Choukri Mohamed Xukri

 duela 2 minutu

English > euskara

8. Bertsioa amaitutakoan argitaratu

Arazo batzuk geratu ohi dira, eta abisuak azter daitezke. Batzuk erraz konpon daitezke batzuetan (konturatu gabeko akatsak adib.) baina beste batzuetan (erreferentziak adib.) hobe da hala ere onartzea eta hurrengo urratsean konpontzea

Publish with existing issues?

Your translation still has potential issues. Please, consider checking the issues before publishing.

Editors reviewing your contribution will appreciate the improvements.

Utzi	Argitaratu halere
-------------	--------------------------

9. Joan orrazketa egitera ohiko Wikipedian.

«Mohamed Xukri» aldatzen - Wikipedia, entziklopedia askea. - Mozilla Firefox

W «Mohamed Xukri» aldatu X +

https://eu.wikipedia.org/wiki/Moh

Inaki.alegria Nire eztabaida Proba orria Hobespenak Beta Jarraipen zerrenda Nire ekarpenak Saioa itxi

Artikuluak Eztabaida +TX Irakurri Aldatu Aldatu iturburu kodea Ikusi historia Gehiago Wikipedia wikian bilatu

Paragrafoa Erreferentzia Txertatu Aldaketak argitaratu...

Mohamed Xukri

Kalitate neurraketa (4.96)

Mohamed Xukri (arabieraz: محمد بكري, amazigeraz: **ⵎⵓⵃⵎⵉⵔ ⵁⵔ ⵁⵔ ⵁⵔ**), Rif aldeko idazle eta nobelagilea izan zen, 1935ko uztailaren 15ean jaioa eta 2003ko azaroaren 15ean hila. Nazioartean ezaguna da bere autobiografia *den Ogi Hutsa (al-Khubz For al-Hafi)* liburu dela-eta, Tennessee Williams dramaturgo estatubatuarak "gizakiaren etsipenaren egiazko dokumentu" gisa deskribatu zuena.

Xukri 1935ean jaio zen, Ayt Chiker (horregatik bere hartutako abizena: **Xukri**), herri txiki bat Rif mendietan, Nador probintziaren barruan, Marokon. Oso familia pobreak hazi zen, eta aita tiraniko bat zuenez ihes egin zuen, eta etxerik gabeko haurra izan zen Tangerren, miseria, prostituzio, indarkeria eta drogez inguratuta. 20 urte zituela irakurtzen eta idazten ikastea erabaki zuen, eta geroago eskolako irakasle bihurtu zen.

1960ko hamarkadan Tanger kosmopolitan Paul Bowles, Jean Genet eta Tennessee Williams ezagutu zituen. Xukriren lehen idazlana *Al-adaq* aldizkarian (Beiruteko hilabetekari bat) argitaratu zen 1966an,^[1] "Al-Unf ala al-shati" ("Indarkeria hondartzan") izenburuarekin. Nazioarteko arrakasta Paul Bowles-en *Al-khoubz Al-Hafi* liburuaren ingelesezko itzulpenarekin lortu zuen 1973an (*For Bread Alone*, Telegram Books). Liburu *Tahar Ben Jelloun*-ek frantsesera itzuli zuen 1980an (Éditions Maspéro), 1982an arabieraz argitaratua eta Marokon zentsuratu zen 1983tik 2000ra. Liburu gero beste 30 hizkuntzara itzuliko litzateke. Euskarara Arantzazu Royok ekarri zuen Igela argitaletxean^[2].

Bere lan nagusiak trilogia autobiografikoa dira, non *Ogi hutsa* lehen liburu den. 1960ko eta 1970eko hamarkadetan ipuin bildumak ere idatzi zituen.

KONTUZ IRUDIEKIN

- Arazo gehiago: egin modu sinplean eta erreparatu argitaratu ondoren

KONTUZ ERREFERENTZIEKIN

- Ez argitaratu bibliografia barneratu gabe. Itzultzeko aukeratu dituzun paragrafoen erreferentziak ez ezabatu. Baina, egun dauden tresnak erreferentziak ondo itzuli/egokitzen ez dituztenez, jatorrizko artikuluan zeuden erreferentzia guztiak azalduko dira euskarazko artikuluan, baina hutsik. Erreferentzia-lista horrelako bat izan daiteke euskarazko artikuluan itzulian:

1. ↑ (Ingelesez) «Operating System Market Share Worldwide» StatCounter Global Stats . Noiz kontsultatua: 2020-06-14.
2. ↑ .
3. ↑ .
4. ↑ .

- Asko galtzea da erreferentziak galtzea Wikipedian. Horrelakoetan jatorrizko artikulutik joan behar da erreferentziaren url-a jasotzen eta url horrekin euskarazkoan automatikoki erreferentzia sartzen. Erreferentziak erraz-erraz txertatuko dituzu url-arekin. Erreferentzia horietako bat eskuz sartzea minutu batzuetako lana izan daiteke. Baina sartu behar duzun erreferentzia bat webean badago segundo gutxi batzuetan egin dezakezu automatikoki ‘ikusizko’ edizioan ‘Aipatu’ botoia eta ‘Automatikoa’ aukeratuta sartu webgunearen helbidea (URLa) eta listo. Beti ez du harrapatzen informazio guztia (jai duzu webguneak metadatu horiek definituta ez baditu, noski), baina sarritan bai, eta oso txukun geratzen da. Berdin gertatzen da erreferentzietako bat liburu bat bada, ISBNa emanda automatikoki lortuko dizkizu liburuaren datu guztiak.

