

Hiztegien Ataria, pibotaje bidez sortutako hiztegiak



Iker Manterola, Iñaki San Vicente, Xabier Saralegi

IEB 2013

2013-05-8

- Hiztegi elebidunak
 - Oinarrizko baliabideak hainbat alorretan
 - Hizkuntzen ikasketa, Itzulpengintza, hizkuntza teknologiak,...
 - Eskuragarritasun arazoa:
 - Sorkuntza garestia
 - Baliabide gutxien duten hizkuntzak zer? Euskara?
 - Irtenbidea:
 - Hiztegiak automatikoki sortu \Rightarrow pibotaje teknikak

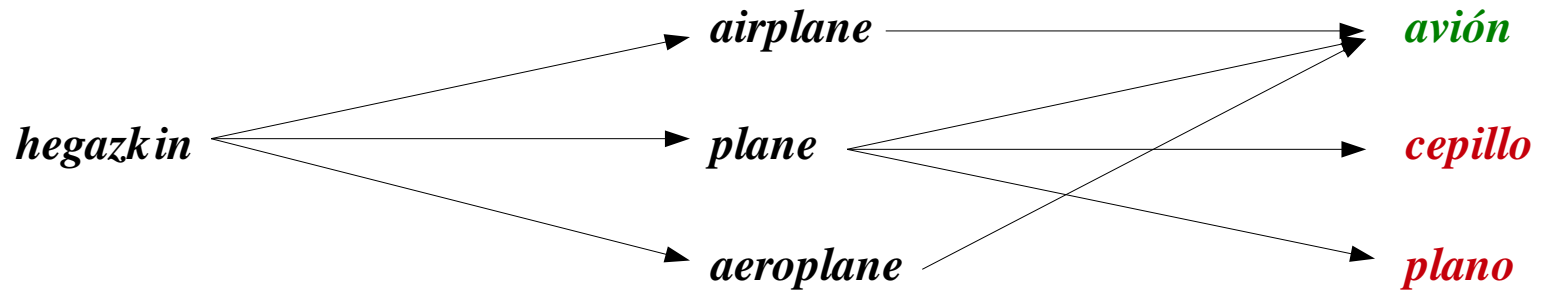
Edukia

- Hiztegien sorrera
- Emaitza hiztegiak
- Hiztegien ataria

- Pibotaje teknikak
 - OINARRIA: Bi hizkuntza erlazionatu hirugarren bat zubi bezala erabiliz:
 - $eu \Rightarrow es + es \Rightarrow en = eu \Rightarrow en$
 - $katu \Rightarrow gato + gato \Rightarrow cat = katu \Rightarrow cat$
 - Zubi hizkuntza baliabide gehien duen hizkuntza
 - ARAZOA: Zarata
 - Itzulpen okerrak sortzen dira polisemiaren eraginez

Hiztegien sorrera (2)

- ZARATA (Polisemia):



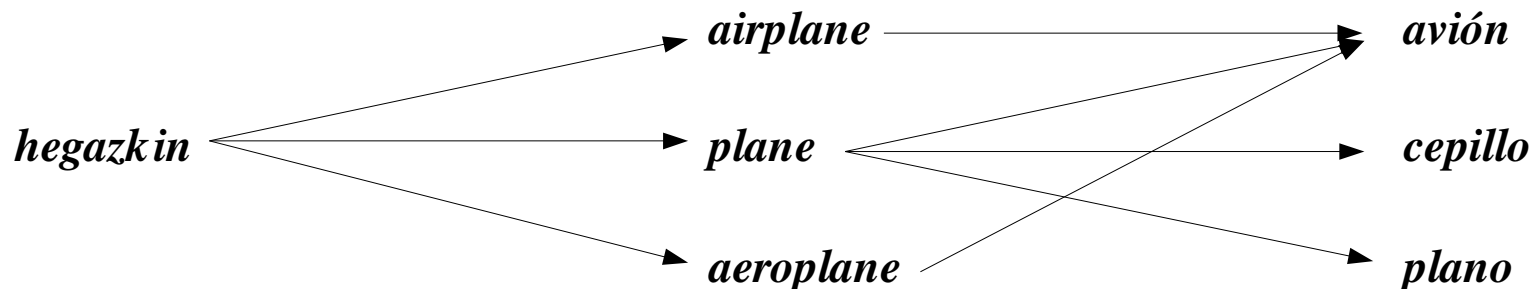
- Zarata gutxitzeko teknikak aplikatzeko beharra

Hiztegien sorrera (3)

- Zarata gutxitzeko aukeraturiko teknikak:
 - *IC (Inverse Consultation)* Tanaka and Umemura, 1994
 - Hiztegien egituraren oinarritua
 - *DS (Distribution Similarity)* Kaji et. al, 2008
 - Hiztegiko sarreran erabileran oinarritua

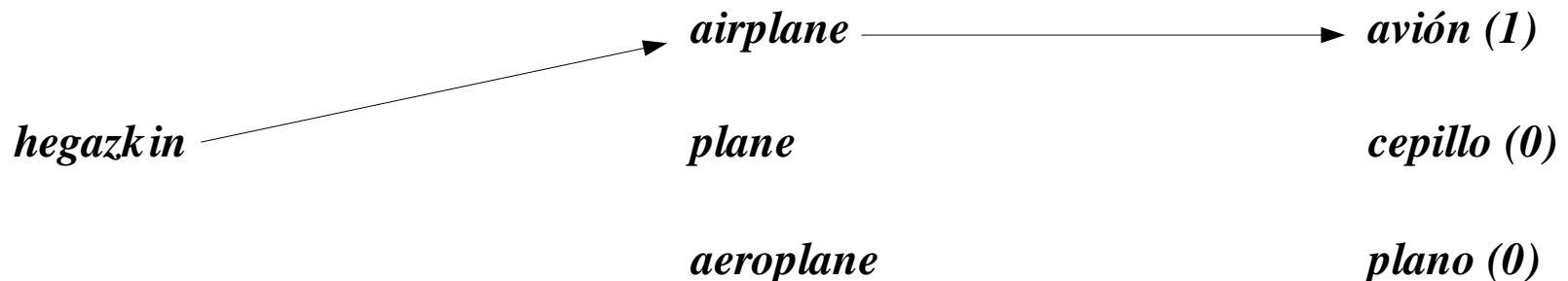
Hiztegien sorrera (4)

- *IC (Inverse Consultation)* teknika:
 - Erabilitako hiztegien egituran oinarritzen da
 - Itzulpenen artean zenbat eta bide gehiago, orduan eta itzulpen probabilitate handiagoa



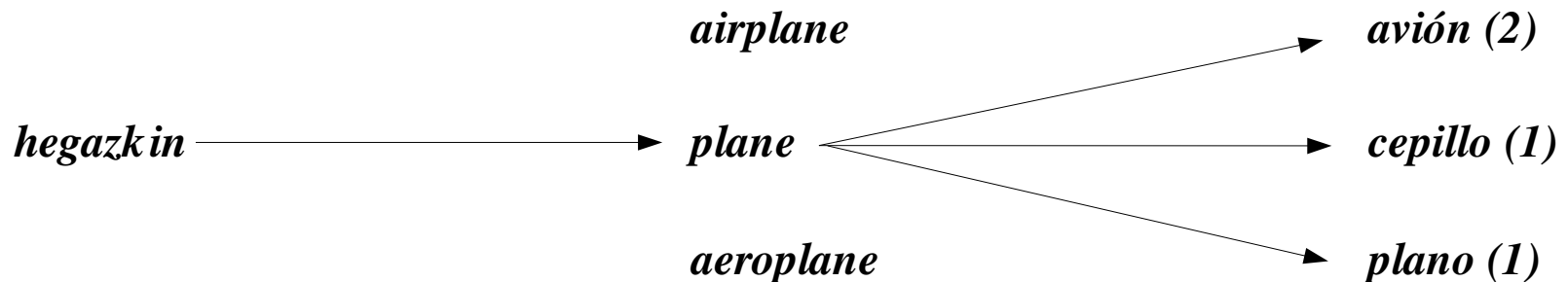
Hiztegien sorrera (4)

- *IC (Inverse Consultation)* teknika:
 - Erabilitako hiztegien egituran oinarritzen da
 - Itzulpenen artean zenbat eta bide gehiago, orduan eta itzulpen probabilitate handiagoa



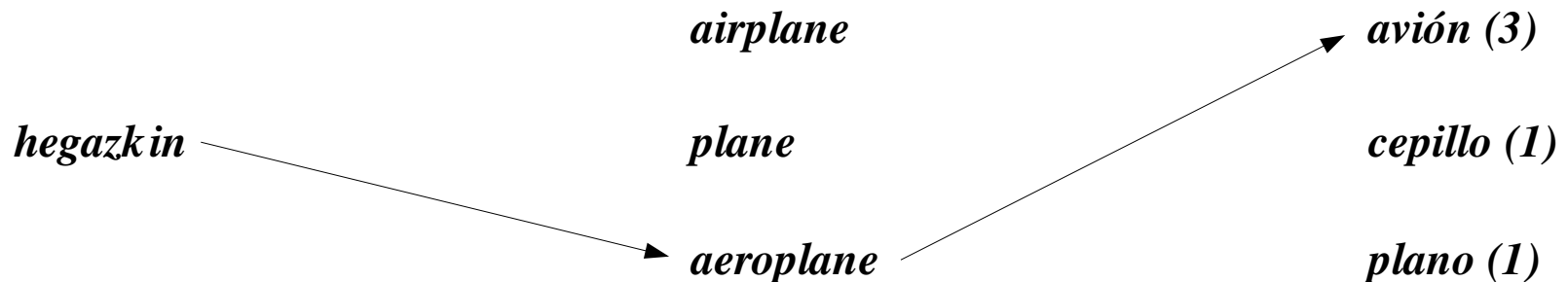
Hiztegien sorrera (4)

- *IC (Inverse Consultation)* teknika:
 - Erabilitako hiztegien egituran oinarritzen da
 - Itzulpenen artean zenbat eta bide gehiago, orduan eta itzulpen probabilitate handiagoa



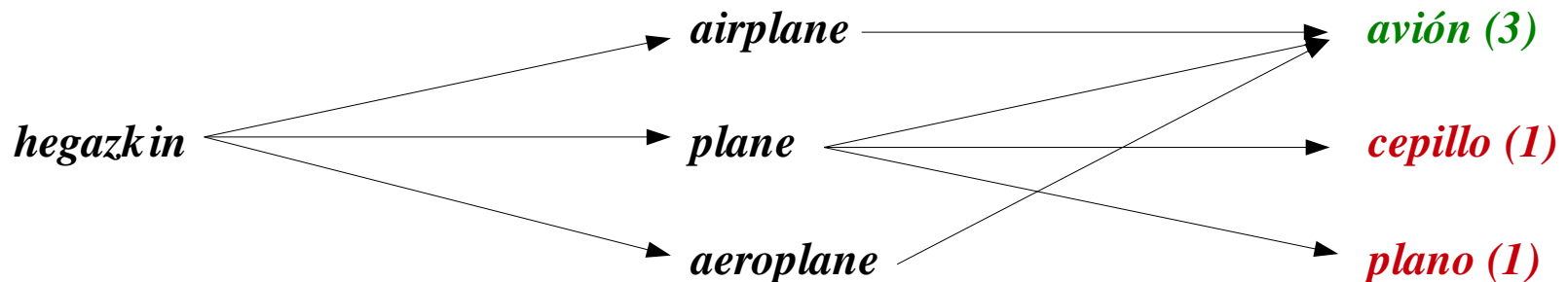
Hiztegien sorrera (4)

- *IC (Inverse Consultation)* teknika:
 - Erabilitako hiztegien egituran oinarritzen da
 - Itzulpenen artean zenbat eta bide gehiago, orduan eta itzulpen probabilitate handiagoa



Hiztegien sorrera (4)

- *IC (Inverse Consultation)* teknika:
 - Erabilitako hiztegien egituran oinarritzen da
 - Itzulpenen artean zenbat eta bide gehiago, orduan eta itzulpen probabilitate handiagoa



Hiztegien sorrera (5)

- *DS (Distributional Similarity)* teknika:
“Hitz jakin baten inguruan agertzen diren hitzak antzekoak dira bi hizkuntzatan”

“Glen Canyon [**urtegia eraiki** denetik, **urteroko** *uholde* **ikaragarriak** eguneroko **marea txiki** batzuk **ordeztu** dituzte; izatez, **erreka** gehienez] lau metro hazten da eskarien arabera ...”

“[The **highest precipitation intensity** and *flooding* **occurred** in the Nisa **river basin** and partly in the Elbe **river basin.**]”

Hiztegien sorrera (6)

- *DS (Distributional Similarity)* teknika:
 - Hitzen testuinguruak testu bilduma elebidunetatik erauzten dira \Rightarrow Corpus Konparagarriak
 - Testuinguru oso antzekoak dituzten itzulpenak zuzentzat hartzen dira

- Automatikoki sortutako 5 hiztegi elebidun (*TEI-XML*):
 - Euskara-Alemana, Euskara-Txinera, Euskara-Hindi, Euskara-Arabiera eta Euskara-Swahili.
 - Zubi hizkuntza: ingelesa
 - Erabilitako hiztegiak:
 - eu-en: Elhuyar hiztegia
 - en-de: *Beolinguus, Wictionary, Wikipedia, Omegawiki*
 - en-zh: *CC-CEDICT, Wictionary, Wikipedia, Omegawiki*
 - en-hi: *Shabdanjali, Wictionary, Wikipedia, Omegawiki*
 - en-ar: *Wictionary, Wikipedia, Omegawiki*
 - en-sw: *Wictionary, Wikipedia, Omegawiki*

Hiztegiak (2)

- Hiztegien tamainak:

	Sarrera kopurua	Itzulpen kopurua
Euskara-Alemana	11.229	31.700
Euskara-Txinera	13.547	83.573
Euskara-Hindi	21.608	57.948
Euskara-Swahili	8.404	12.767
Euskara-Arabiera	11.487	26.377

Hiztegien ataria

- Hiztegien Ataria, pibotaje bidez sortutako hiztegien kontsulta



- 3 atal nagusi:
 - Bilaketa
 - Deskargak
 - Eztabaidagunea

Hiztegien ataria (2)

- Bilaketa: Bilaketa tresna

Hiztegi Kontsulta

Euskara→Alemana garagardo Bilatu

Emaita: 1 sarrera topatu dira hiztegian.

Topatutako sarrerak	Ordainak
<p>garagardo</p> <ul style="list-style-type: none">Lendakari's Muertosko kideekin ez dakit nola moldatuko nintzatekeen ostatu baten barnean garagardoka.Gustatzen zait basoa garagardoz bete berria duenari dena isurtzen zaionean.Txekiarrek sartu zuten Vietnamen garagardo hori, eta hamar dolar ordainduta, ehun bat eros daitezke eta lagun onak egin. <p>Erakutsi adibide gehiago</p>	<p>1 itzulpen ziur eta 2 zalantzako aurkitu dira Erakutsi zalantzakoak</p> <p><u>Bier</u> +0 -0 </p> <ul style="list-style-type: none">Sie mögen Gartenzwerge, trinken Bier und verbringen den Sommer dauergrillend: Dieses Klischee des Wohnwagenbesitzers hat sich spätestens mit der Comedyserie Die Camper verbreitet.Die Souvenirhändler langweilen sich vor verstaubten Teppichen, im Bistro erklärt der Kellner einem enttäuschten Gast, dass es im Hotel kein Bier mehr gebe.Wenn Nelson eine Frau gefällt, bringt sein aufgeblasener Hals einen sonoren Ton hervor, für den andere Männer eine leere Bierflasche brauchen. <p>Erakutsi adibide gehiago</p> <p><u>Gerstensaft</u> +0 -0 </p> <ul style="list-style-type: none">Es sind die beiden Hauptprodukte des beliebtesten Bierparadieses Österreichs, in der die Gäste an heißen Tagen so viel Gerstensaft in sich hineinschütten wie nirgendwo anders im Land.Der Bierbrauer, der erkannt hat, dass seine Kunden nicht irgendeinen Gerstensaft, sondern "veräffelte Heimat" trinken wollen, kann die entsprechenden Werbe-Bilder und Botschaften liefern, so den Umsatz erhöhen, den Marktforscher bezahlen und den Kunden glücklich machen.Auch die Biertrinker im Ausland griffen verstärkt zum deutschen Gerstensaft - die Exporte legten um 12,6 Prozent zu. <p>Erakutsi adibide gehiago</p> <p><u>Molle</u> +0 -0 </p> <ul style="list-style-type: none">"In Schintoismus und Buddhismus haben auch die Dinge eine Seele", erklärt die Heidelberger Altersforscherin Heidrun Mollenkopf die ostasiatische Technikbegeisterung."Die einzige Möglichkeit, heutzutage eine vernünftige ökologische Balance zu erhalten, besteht darin, die Menschen zu verbannen - vollständig", räsoniert Millionär Stef Moller in seinem Privatreservat - keineswegs die Meinung eines isolierten Spinners.Zwar kann der sozialistische Premierminister Guy Mollet die Abstimmung in Frankreich nur mit Mühe, und nur weil die deutschen Sozialdemokraten schon zugestimmt haben, gewinnen. <p>Erakutsi adibide gehiago</p>

Hiztegien ataria (2)

- Bilaketa: Itzulpenak

Hiztegi Kontsulta

Euskara→Alemana | garagardo | Bilatu

Emaita: 1 sarrera topatu dira hiztegian.

Topatutako sarrerak	Ordainak
<p>garagardo</p> <p>■ ...ndakari... bertosko kideekin ez dakit nola moldatuko nintzatekeen ostatu baten barnean garagardoka.</p> <ul style="list-style-type: none">■ Gustatzen zait basoa garagardoz bete berria duenari dena isurtzen zaionean.■ Txekiarrek sartu zuten Vietnamen garagardo hori, eta hamar dolar ordainduta, ehun bat eros daitezke eta lagun onak egin. <p>Erakutsi adibide gehiago</p>	<p>1 itzulpen ziur eta 2 zalantzako aurkitu dira Erakutsi zalantzakoak</p> <p>Bier +0 -0 </p> <p>■ ...mögen Gartenzwerge, trinken Bier und verbringen den Sommer dauergrillend: Dieses Klischee des Wohnwagenbesitzers hat sich spätestens mit der Comedyserie Die Camper verbreitet.</p> <ul style="list-style-type: none">■ Die Souvenirhändler langweilen sich vor verstaubten Teppichen, im Bistro erklärt der Kellner einem enttäuschten Gast, dass es im Hotel kein Bier mehr gebe.■ Wenn Nelson eine Frau gefällt, bringt sein aufgeblasener Hals einen sonoren Ton hervor, für den andere Männer eine leere Bierflasche brauchen. <p>Erakutsi adibide gehiago</p> <p>Gerstensaft +0 -0 </p> <p>■ ...sind die ...den Hauptprodukte des beliebtesten Bierparadieses Österreichs, in der die Gäste an heißen Tagen so viel Gerstensaft in sich hineinschütten wie nirgendwo anders im Land.</p> <ul style="list-style-type: none">■ Der Bierbrauer, der erkannt hat, dass seine Kunden nicht irgendeinen Gerstensaft, sondern "versäffete Heimat" trinken wollen, kann die entsprechenden Werbe-Bilder und Botschaften liefern, so den Umsatz erhöhen, den Marktforscher bezahlen und den Kunden glücklich machen.■ Auch die Biertrinker im Ausland griffen verstärkt zum deutschen Gerstensaft - die Exporte legten um 12,6 Prozent zu. <p>Erakutsi adibide gehiago</p> <p>Molle +0 -0 </p> <p>■ ...hintoismus und Buddhismus haben auch die Dinge eine Seele", erklärt die Heidelberger Altersforscherin Heidrun Mollenkopf die ostasiatische Technikbegeisterung.</p> <ul style="list-style-type: none">■ "Die einzige Möglichkeit, heutzutage eine vernünftige ökologische Balance zu erhalten, besteht darin, die Menschen zu verbannen - vollständig", räsoniert Millionär Stef Moller in seinem Privatreservat - keineswegs die Meinung eines isolierten Spinners.■ Zwar kann der sozialistische Premierminister Guy Mollet die Abstimmung in Frankreich nur mit Mühe, und nur weil die deutschen Sozialdemokraten schon zugestimmt haben, gewinnen. <p>Erakutsi adibide gehiago</p>

Hiztegien ataria (2)

- Bilaketa: Adibideak

Hiztegi Kontsulta

Euskara→Alemana

Emitza: 1 sarrera topatu dira hiztegian.

Topatutako sarrerak	Ordainak
<p>garagardo</p> <ul style="list-style-type: none">Lendakari's Muertosko kideekin ez dakit nola moldatuko nintzatekeen ostatu baten barnean garagardoka.Gustatzen zait basoa garagardoz bete berria duenari dena isurtzen zaionean.Txekiarrek sartu zuten Vietnamen garagardo hori, eta hamar dolar ordainduta, ehun bat eros daitezke eta lagun onak egin. <p>Erakutsi adibide gehiago</p>	<p>1 itzulpen ziur eta 2 zalantzako aurkitu dira Erakutsi zalantzakoak</p> <p>Bier +0 <input type="button" value="👍"/> -0 <input type="button" value="👎"/></p> <ul style="list-style-type: none">Sie mögen Gartenzwerge, trinken Bier und verbringen den Sommer dauergrillend: Dieses Klischee des Wohnwagenbesitzers hat sich spätestens mit der Comedyserie Die Camper verbreitet.Die Souvenirhändler langweilen sich vor verstaubten Teppichen, im Bistro erklärt der Kellner einem enttäuschten Gast, dass es im Hotel kein Bier mehr gebe.Wenn Nelson eine Frau gefällt, bringt sein aufgeblasener Hals einen sonoren Ton hervor, für den andere Männer eine leere Bierflasche brauchen. <p>Erakutsi adibide gehiago</p> <p>Gerstensaft +0 <input type="button" value="👍"/> -0 <input type="button" value="👎"/></p> <ul style="list-style-type: none">Es sind die beiden Hauptprodukte des beliebtesten Bierparadieses Österreichs, in der die Gäste an heißen Tagen so viel Gerstensaft in sich hineinschütten wie nirgendwo anders im Land.Der Bierbrauer, der erkannt hat, dass seine Kunden nicht irgendeinen Gerstensaft, sondern "verätselte Heimat" trinken wollen, kann die entsprechenden Werbe-Bilder und Botschaften liefern, so den Umsatz erhöhen, den Marktforscher bezahlen und den Kunden glücklich machen.Auch die Biertrinker im Ausland griffen verstärkt zum deutschen Gerstensaft - die Exporte legten um 12,6 Prozent zu. <p>Erakutsi adibide gehiago</p> <p>Molle +0 <input type="button" value="👍"/> -0 <input type="button" value="👎"/></p> <ul style="list-style-type: none">"In Schintoismus und Buddhismus haben auch die Dinge eine Seele", erklärt die Heidelberger Altersforscherin Heidrun Mollenkopf die ostasiatische Technikbegeisterung."Die einzige Möglichkeit, heutzutage eine vernünftige ökologische Balance zu erhalten, besteht darin, die Menschen zu verbannen - vollständig", räsoniert Millionär Stef Moller in seinem Privatreservat - keineswegs die Meinung eines isolierten Spinners.Zwar kann der sozialistische Premierminister Guy Mollet die Abstimmung in Frankreich nur mit Mühe, und nur weil die deutschen Sozialdemokraten schon zugestimmt haben, gewinnen. <p>Erakutsi adibide gehiago</p>

Hiztegien ataria (2)

- Bilaketa: Boto sistema

Hiztegi Kontsulta

Euskara→Alemana garagardo Bilatu

Eraitza: 1 sarrera topatu dira hiztegian.

Topatutako sarrerak	Ordainak
<p>garagardo</p> <ul style="list-style-type: none">Lendakari's Muertosko kideekin ez dakit nola moldatuko nintzatekeen ostatu baten barnean garagardoka.Gustatzen zait basoa garagardoz bete berria duenari dena isurtzen zaionean.Txekiarrek sartu zuten Vietnamen garagardo hori, eta hamar dolar ordainduta, ehun bat eros daitezke eta lagun onak egin. <p>Erakutsi adibide gehiago</p>	<p>1 itzulpen ziur eta 2 zalantzako aurkitu dira</p> <p>Bier</p> <p>+0 -0</p> <ul style="list-style-type: none">Sie mögen Gartenzwerge, trinken Bier und verbringen den Sommer dauegerillend; Dieses Klischee des Wohnwagenbesitzers hat sich spätestens mit der Comedyserie Die Camper verbreitet.Die Souvenirhändler langweilen sich vor verstaubten Teppichen, im Bistro erklärt der Kellner einem enttäuschten Gast, dass es im Hotel kein Bier mehr gebe.Wenn Nelson eine Frau gefällt, bringt sein aufgeblasener Hals einen sonoren Ton hervor, für den andere Männer eine leere Bierflasche brauchen. <p>Erakutsi adibide gehiago</p> <p>Gerstensaft</p> <p>+0 -0</p> <ul style="list-style-type: none">Es sind die beiden Hauptprodukte des beliebtesten Bierparadieses Österreichs, in der die Gäste an heißen Tagen so viel Gerstensaft in sich hineinschütten wie nirgendwo anders.Der Bierbrauer, der erkannt hat, dass seine Kunden nicht irgendeinen Gerstensaft, sondern "verätselte Heimat" trinken wollen, kann die entsprechenden Werbe-Bilder und Botschaften liefern, so den Umsatz erhöhen, den Marktforscher bezahlen und den Kunden glücklich machen.Auch die Biertrinker im Ausland griffen verstärkt zum deutschen Gerstensaft - die Exporte legten um 12,6 Prozent zu. <p>Erakutsi adibide gehiago</p> <p>Molle</p> <p>+0 -0</p> <ul style="list-style-type: none">"In Schintoismus und Buddhismus haben auch die Dinge eine Seele", erklärt die Heidelberger Altersforscherin Heidrun Mollenkopf die ostasiatische Technikbegeisterung."Die einzige Möglichkeit, heutzutage eine vernünftige ökologische Balance zu erhalten, besteht darin, die Menschen zu verbannen - vollständig", räsoniert Millionär Stef Moller in seinem Privatreservat - keineswegs die Meinung eines isolierten Spinners.Zwar kann der sozialistische Premierminister Guy Mollet die Abstimmung in Frankreich nur mit Mühe, und nur weil die deutschen Sozialdemokraten schon zugestimmt haben, gewinnen. <p>Erakutsi adibide gehiago</p>

Hiztegien ataria (3)

- Deskargak:
 - Hiztegiak bi norabidetan eskaintzen dira
 - TEI (XML) estandarrean kodetuak
 - Itzulpen ziurrak bakarrik esportatzen dira

Hiztegien Deskarga


Hiztegien atarian kontsultagarri dauden hiztegi guztiak deskargarri daude hemen:

Euskara - Alemana	Alemana - Euskara
Euskara - Txinera	Txinera - Euskara
Euskara - Hindi	Hindi - Euskara
Euskara - Swahili	Swahili - Euskara
Euskara - Arabiera	Arabiera - Euskara

Hiztegien ataria (4)









- Eztabaidagunea

FAQ Search The team Register Login

 **Eztabaidagunea**
Elhuyar I+G

Jump to: Board index

View unanswered posts • View active topics

HIZTEGIEN ATARIA - EZTABAIDAGUNEA		STATISTICS	LAST POST
	Webguneari buruzko Zalantzak/Iradokizunak	Topics: 1 Posts: 1	 Fri Jan 25, 2013 12:40 pm by HiztegienAtaria
	EU - AR Euskara - Arabiera zalantzak / الباسك - العربية الشكوك	Topics: 0 Posts: 0	No posts
	EU - DE Euskara - Alemana zalantzak / Baskisch - Deutsch Zweifel	Topics: 1 Posts: 3	 Fri Feb 01, 2013 12:44 pm by isanvicente
	EU - HI Euskara - Hindi zalantzak / बास्क - हिन्दी संदेह	Topics: 0 Posts: 0	No posts
	EU - SW Euskara - Swahili zalantzak / Basque - Kiswahili mashaka	Topics: 0 Posts: 0	No posts
	EU - ZH Euskara - Txinera zalantzak / 巴斯克 - 中国疑虑	Topics: 0 Posts: 0	No posts

LOGIN • REGISTER

Username: Password: | Log me on automatically each visit

Hiztegien ataria (5)

- Etorkizuneko helburuak:
 - Hiztegiak modu kolaboratiboan hobetu
 - Erabilitako teknikak hobetu
 - Hizkuntza bikote gehiago gehitu
 - Zuzenketa aukerak areagotu:
 - Itzulpenak aldatu, gehitu edo kentzeko aukera
 - Adibideak editatzeko aukera

- Erreferentziak:
 - X. Saralegi, I. Manterola, I. San Vicente. 2011. “Analizing Methods for Improving Precision of Pivot Based Bilingual Dictionaries”. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*. Edinburgo.
 - X. Saralegi, I. Manterola, I. San Vicente. 2012. "Building a Basque-Chinese Dictionary by using English as a Pivot". *In Proceedings of the 8th international conference on Language Resources and Evaluation, LREC'12*. 23-25 May, Istanbul, Turkia.

Eskerrik asko!

Probatu eta erabili!

<http://hiztegieleaniztunak.elhuyar.org/>