

Taligner: itzulpen corpus eleaniztunak sortzeko tresna

IBON URIBARRI ZENEKORTA

Euskal Herriko Unibertsitatea (UPV/EHU)

Artikulu honetan, Taligner programa informatikoa aurkeztuko dut. Software horrek corpus paraleloak sortzeko, kudeatzeko eta analizatzeko balio du, corpus eleaniztunak eta itzulpen corpusak barne. Programaren jatorria eta ezaugarri nagusiak zehaztuko ditut, bereziki.

Jatorria

EHUko Itzulpen eta Interpretazioko jakintza-arloko ikerketa ekimenetan du jatorria Taligner programak. Bai TRACE ikerketa proiektuetan bai Gasteizen egin diren zenbait doktoretza tesitan, corpusak erabili dira zenbait itzulpen errealitate hobeto ezagutzeko. Hasiera batean, corpus horien kudeaketa oso oinarrizkoa izan zen: Word programarekin lerrokatzen ziren jatorrizko testua eta itzulpena(k). Egoera horrek ez gintuen asetzen, eta ordurako eskuragarri zeuden zenbait software probatu genituen. Jauzi nabarmena egin zen Elizabete Manterolaren tesiari esker. Batetik, itzulpen corpus paralelo eleaniztun zabala sortu zuen Manterolak (Atxagaren 12 testu oso edo lagin, eta haien itzulpenak 7 hizkuntzatara; milioi eta erdi hitz inguru guztira), eta, horretarako, Wordsmith Tools programa erabili zuen. Tresna hori zen une hartan aurkitu genuen onena; izan ere, corpusak analizatzeko tresneria garatuarekin uztartzen du corpus paraleloak sortzeko tresna. Hala ere, zenbait arazo edo oztopo topatu genituen: batetik, software ez-librea erabiltzen du; bestetik, artxibo mota propio ez-estandarrak baliatzen ditu, eta, batez ere, lerrokatzeko tresneria oso astuna du, hau da, jatorrizko testua eta itzulpena(k) esaldi mailan doitzeko prozesua oso neketsua da.



i. irudia: Atxagaren itzulpen corpusa Wordsmith Tools erabiliz.

Hala, ikusi genuen aurrera egin nahi bagenuen geure beharretara egokitutako tresna bat sortzea zela bidea. Une horretan, gainera, Hezkuntza Ministerioko TRACE proiektua lortu genuen, Leóngo Unibertsitateko talde batekin elkarlanean. Orduan, Leónen egin ziren Trace aligner i.o. sortzeko lehen urratsak. Bitestuak lerrotzeko oinarritzko programa bat sortu zen, baina oso erabilera mugatua zuen: lerrotutako testuak (laburrak) html formatuan erakustera mugatzen zen.

Gerxeago, Gasteizen, Iñaki Albisua informatikaria kontratatzeko aukera izan genuen, TRACE proiektuen diru laguntzei esker hasieran eta TRALIMA/ITZULIK ikerketa taldeko baliabideei esker ondoren, eta horrek ekarri zuen TRACEaligner 2.o. garatzeko bultzada erabakigarria. Leónen egin zuten oinarri hartuta, Albisuak programa garatu zuen, bi doktoretza tesiren eskakizunak betetzeko, Naroa Zubillagaren tesirako lehenik eta Zuriñe Sanzen tesirako ondoren: programaren aldaera berriak testu luzeekin lan egiteko aukera ematen zuen; tritestuak lerrotzeko bidea ireki zuen (jatorrizko testu alemanez eta euskarazko itzulpenez gain tarteko gaztelaniazko testua ere behar baikenuen zeharkako itzulpenen kasuan); lerroketa doitzeko funtzioak gaineratu ziren; corpusak mysql formatuko datu base bihurtzen zituen, eta bilaketa motorra garatu zen sortutako corpusa arakatzeko. Une horretan auzo nagusia datu basea sortzeko prozesu astuna zen (xampp programa erabili behar genuen sql datu baseak sortzeko).

Programaren aldaera hori Aleuska egitasmoko zenbait lani esker eta, aldi berean, lan horien mesedetarako garatu zen. Alemanetik euskarara itzulitako testuen katalogoan oinarrituta (700 sarreratik gora), Naroa Zubillagak AleuskaHGL itzulpen corpus paraleloa sortu zuen (18 egileren 19 itzulpen zuzen eta 14 zeharkako; 1.275.000 hitz guztira); Zuriñe Sanzek, AleuskaPhraseo corpora (30 egileren 34 itzulpen zuzen eta 14 zeharkako; 3.500.000 hitz guztira), eta Ibon Uribarrik, AleuskaFilo corpora (13 egileren 33 itzulpen zuzen; 1.200.000 hitz guztira). Aleuska corpusek gaur-gaurkoz 5.500.000 hitz inguru biltzen dituzte osotara. Bestek beste, Touryren interferentziaren legea zehazteko balio izan du. Touryk zeharkako interferentzia aipatzen du (1995: 72):

Hypothetically identified relationships may also give rise to the assumptions that a target text drew on a text in a language other than the assumed one, or on more than one source text, in more than one language.

Baina, gero, ez du zehazten interferentziaren legeaz diharduenean. Izan ere, paratestuen mailan itzulpen zuzen moduan aurkezten diren testuetan, zeharkakotasun arrastoak nabari dira Aleuska corpusetan, zenbait mailatan. Gerta daiteke testua itzulpen zuzen moduan aurkeztea baina benetan gaztelaniatik itzulia izatea testua, ez alemanetik, hau da, ustez itzulpen zuzena zena (*assumed direct translation*) ezkutuko zeharkako itzulpena izatea (*covert indirect translation*). Testu interferentzia hori osoa izan daiteke, jatorrizko testua erabili ez denean, edo gertakari bakana izan daiteke, aldian-aldian jatorrizko testuaz gain bitarteko itzulpena(k) erabili d(ir)elako (itzulpen konpilatua litzateke hori). Zeharkako testu interferentzia horretaz gain, zeharkako interferentzia instrumentala ere identifikatu dugu, alemanetik euskarara itzultzerakoan zeharkako laguntza tresnak erabiltzen direlako (hiztegiak, bereziki). Eta, azkenik, zeharkako gogo interferentzia edo interferentzia kognitiboa ere aurkitu dugu, hau da, alemanetik zuzenean itzultzerakoan, euskal itzultzaileak zeharkako hizkuntza bat ere aktibatzen du (gaztelania, bereziki), eta, orduan, zeharkako itzulpena egiteko aukera sortzen da noizean behin, tarteka edo sarriago. Beraz, itzulpen zuzena/zeharkakoa dikotomia oso motz geratzen da itzulpen errealitate askoz ere konplexuago bat aztertzeko. Azaldutako itzulpen egoera horri *zeharkako interferentzia diglosikoa* deitu diogu. Ustez (guztiz) jatorrizkoak diren testu produkzioan gertatzen den fenomeno bera da, azken finean.

Azken TRACE proiektuari esker (EHU, Leóngo Unibertsitatea, Kantabriako Unibertsitatea, Elkeko Unibertsitatea), Taligner 3.0. garatzeko aukera ere izan dugu. Iñaki Albiñak hutsetik berridatzi du programa osoa. Oraintsu arte, adabakiz osatutako kode nahasi bat genuen, eta arazo teknikoak sortzen ziren, askotan. Orain, java hizkuntzan idatzitako programa multiplataforma bat dugu, hau da, java behar du ordenagailuak programa erabiltzeko, baina berdin dabil, printzipioz, Windows, Mac eta Linux plataformetan. Hauek dira, oraingoz, azkenen den aldaera horren berritasun nagusiak: pluritestuz osatutako corpusak osatzeko aukera ematen du, hiru testuko muga gaindituz; testu dramatikoaz osatutako corpusak sortzeko aukera berezitua ematen du; corpusaren testu analisia egiteko lehen tresnak biltzen ditu (testu hustuketa egiteko tresna); datu baseak sortzeko tresna integratu da (orain ez da xampp edo beste

programa berezirik behar); hainbat corpus aldi berean kudeatzeko funtzioa gaineratu zaio. Al- daera hori Itzulpen Ikasketetako gradu amaierako lan batzuk egiteko baliatu da; Itzulpena eta Teknologia graduondokoa ere erabili da, eta Garazi Arrulak 10 idazleren 10 testu eta haien autoitzulpenak biltzen dituen corpora osatzeko baliatu du.

Programaren ezaugarri nagusiak

Hasiera batean poesia testuak ere lantzea pentsatu bazen ere, oraingoz, aukera hori ez da garatu, eta testu laua eta testu dramatikoak erabiltzeko dago prestatuta programa. Hasieran, testu laua bakarrik hartu genuen kontuan, eta, azken aldaeran, testu dramatikoak gaineratu diogu, hau da, elkarrizketa testuaren egitura konplexuagoa lantzeko aukera jaso dugu: pertsonaien izenak, pertsonaien solasa, iruzkinak.



2. irudia: Taligner 3.0. interfazea.

Orain, funtzio nagusiak aurkeztuko ditut, ezkerretik eskuinera. Lehenak *Limpiar* izena du (oraingoz, gaztelaniaz bakarrik dugu interfazea), eta testu garbiketan laguntzen duen eranskin bat dela esan daiteke. Hainbatetan, testuak eskaneatu egin behar izan ditugu, eta horrek testuak garbitzeko lan astuna ekartzen zuen gero. Testu garbiketa horretako eragiketa errepikakor batzuk kolpe batean egiteko tresna bat sortu zuen Albiuak, Access erabiliz, eta tresna auto-

moa zena programa nagusian integratu genuen, azkenean. Baina, zuzenean testu digitalarekin lan egiteko aukera izanez gero, funtzio hori ez da ezertarako behar.

Etiketazioa da bigarren funtzioa. Lerrokatzen hasi aurretik, testuak etiketatu egin behar dira. Abiapuntu gisa, txt formatuko testuak erabiliko ditugu beti, eta, etiketatze funtzio honekin, xml artxiboak lortuko ditugu. Testua aukeratu eta kargatu ondoren, *Etiquetar texto* botoia erabiliz, behin-behineko emaitza ikusiko dugu, eskuinean.



3. irudia: Testuak etiketatzen. Canettiren *Die gerettete Zunge*.

Gure kasuan, testu bat etiketatuta egoteak esan nahi du, alde batetik, paragrafo eta esaldiak etiketa (*tag*) bidez markatuta egongo direla, eta, bestetik, testuak metadatu hornituko ditugula; gero bilaketak egiteko ezinbestekoa izango den informazioz, alegia. Etiketa horretan, testuaren izenburua, egilea, itzultzailea, eta itzulpen eta genero mota jasotzen dira, besteak beste.

Ondoren, sortuko den fitxategi berriari izen bat jarri, eta gorde egin beharko dugu (*Guardar texto*), eta, hala, xml fitxategi bat sortuko da; hor amaitzen da etiketatze prozesua. Eragiketa bera egin behar dugu corpusean jaso nahi dugun testu ororekin. Xml fitxategiak software jakin bati lotuak ez daudenez, testu kudeatzaile arrunt bat erabil dezakegu fitxategiaren barneinformazioa ikusteko.



4. irudia: Metadatuak. Canettiren *Die gerettete Zunge*.



5. irudia: Xml dokumentua. Kafkaren *Die Verwandlung*.

Irudian, fitxategiaren zati bat besterik ez da ikusten. Garrantzitsuena bi zati handi daudela identifikatzea da: alde batetik, <header> etiketaren barnean biltzen dena, eta, bestetik, <text> barruan dagoena. Lehen zati horretan, metadatuak aurkitzen dira (izenburua, egilea, kodea, itzultzailea, hizkuntza, modua, generoa), eta bigarrean, aldiz, testua bera dago. Testua, era berean, paragrafo eta esalditan banatuta dago. Paragrafo eta esaldi gisa markatzeko, berariazko etiketak erabiltzen dira: (<p> eta <s>). Goiko irudian gorritz agertzen den *n*-a paragrafo eta esaldien atributu bat da, eta halakoak zenbatzeko balio du. Beheko eskeman ikus daiteke aipatu berri duguna, laburbilduta. Sarean erabiltzen ditugun corpusek ere egitura bera dute (TEI estandarra).



6. irudia: Xml testu baten egitura.

Testuak etiketatu ondoren, lerrotatu egin beharko ditugu, hirugarren funtzio nagusia erabiliz, *Alinear*. Bi, hiru edo testu gehiago lerrotatu daitezke aldi berean, aurretik sortu ditugun xml artxiboetatik abiatuz. Segundo gutxiren buruan lerrotatuko du programak testu bikote, hirukote, laukotea... (testuak oso luzeak baldin badira, denbora pixka bat har dezake kontuak). 7. irudian, behin testuak lerrotatuta programaren itxura ikus dezakegu.

Hala ere, gehienetan, lerrotatze-prozesu automatikoa ez da guttiz zehatza izango, are guxtiago literatur itzulpenak lerrotatu nahi baditugu, hau da, programak testu bakoitza esalditan banatzen du, eta testuak esaldiz esaldi uztartzen, baina uztarketa hori ez da *zuzena* izango beti; bereziki, gorabeherak izan badira itzulpen prozesuan puntuazioarekin. Orduan, guri egokituko zaigu, jatorrizko testuan oinarrituta, jatorrizko testuko esaldia xede testuko esaldiarekin bat datorrela ziurtatzea.

Active		
		
		
		
		
		
		
UR-130 ur - length 107	UR-130 ur - length 107	UR-062 ur - length 1447
1# URLO	2# Resurrección M	1# URLO
	3# Ardon	2# Resurrección M
	4# URLO	3# Ardon
		4# URLO
UR-130 ur - length 107	UR-130 ur - length 107	UR-062 ur - length 1447
1# URLO	2# Resurrección M	1# URLO
	3# Ardon	2# Resurrección M
	4# URLO	3# Ardon
		4# URLO
UR-130 ur - length 107	UR-130 ur - length 107	UR-062 ur - length 1447
1# URLO	2# Resurrección M	1# URLO
	3# Ardon	2# Resurrección M
	4# URLO	3# Ardon
		4# URLO
UR-130 ur - length 107	UR-130 ur - length 107	UR-062 ur - length 1447
1# URLO	2# Resurrección M	1# URLO
	3# Ardon	2# Resurrección M
	4# URLO	3# Ardon
		4# URLO
UR-130 ur - length 107	UR-130 ur - length 107	UR-062 ur - length 1447
1# URLO	2# Resurrección M	1# URLO
	3# Ardon	2# Resurrección M
	4# URLO	3# Ardon
		4# URLO
UR-130 ur - length 107	UR-130 ur - length 107	UR-062 ur - length 1447
1# URLO	2# Resurrección M	1# URLO
	3# Ardon	2# Resurrección M
	4# URLO	3# Ardon
		4# URLO
UR-130 ur - length 107	UR-130 ur - length 107	UR-062 ur - length 1447
1# URLO	2# Resurrección M	1# URLO
	3# Ardon	2# Resurrección M
	4# URLO	3# Ardon
		4# URLO
UR-130 ur - length 107	UR-130 ur - length 107	UR-062 ur - length 1447
1# URLO	2# Resurrección M	1# URLO
	3# Ardon	2# Resurrección M
	4# URLO	3# Ardon
		4# URLO

7. irudia: R. M. Azkueren *Urlo* opera, euskaraz, gaztelaniaz eta alemanez.

Active		
		
		
		
		
		
		
ComedienGek ur - length 698	ComedienGek ur - length 698	ComedienGek ur - length 698
1# RUSTOSK 1945-1911	1# RUSTOSK 1945-1911	1# RUSTOSK 1945-1911
2# Meine liebste Erinnerung	2# My Dearest Memory	2# Mi recuerdo más querido
3# Meine liebste Erinnerung ist in Rot gezeichnet	3# My dearest memory is signed in red	3# Mi recuerdo más querido está sellado de rojo
4# Auf dem Arm eines Mahlkorns kommt ich zu dem Türhüter, der Roten vor mir ist, und zur Linken geht eine Treppe hinunter, die	4# I come out of a door on the arm of a mill, the door is lined with red, and to the left a staircase goes down, equally red.	4# Salgo por una puerta en brazos de una muela, sobre mí el suelo es rojo y a la izquierda descendiendo una escalera guarnecida roja.
5# Opposider mir was, er schief über, schief sich aber Tiere und ein	5# Across from us, at the same height, a deer comes, and a ending	5# Frente a nosotros, a la misma altura, se abre una puerta y aparece un
6# Es soll ganz schön an euch herab, nicht stehen und sagt es mir	6# He stays right up close	6# De su apertura resulta, se detiene, y me dice
7# "Gag de Zagep"	7# "There was your tongue"	7# "¡Esa es la lengua!"
8# Ich werde die Zage herab, er geht in seine Tasche, und ein	8# I stick out my tongue, he shrugs, opens it, and he	8# Tu boca la lengua, él palpa en su bolsillo, extrae una lengua, la abre y
9# Er sagt "Viel schmecken wir das de Zage-ah!"	9# He says "How well it	9# dice "¡Bueno la confitección la lengua!"
10# Ich sage er sagt, die Zage zurückzubringen, er kommt wieder	10# I don't dare pull back my tongue, so return center and close, the	10# Yo me atrevo a retirar la lengua, él se acerca cada vez más hasta
11# In diese Augenblick sagt er die Meiner zurück, sagt	11# In the last moment, he pulls back the leaf, saying	11# En el último momento retira la lengua y dice
12# "Nicht noch mal, morgen!"	12# "No today, tomorrow!"	12# "No volverás ya, mañana!"

8. irudia: Lerrokatze funtzioaren doikuntza tresnak.

Itzulpena(k) jatorrizkoari egokitzeak, beharrezko aldaketak egin ahal izateko, goiko irudian ikus daitezkeen botoiak ditugu (saguaren eskuinaldean sakatuta ikusiko ditugu). *Combinar* botoiarekin, bi esaldi batu ditzakegu; laukiak gehitu nahi baditugu, *Insertar blanco*-n egingo dugu klik; hitzak berak dioen bezala, *Dividir*-ekin, laukiak zatitu daitezke; *Eliminar*-en klik eginez, nahi beste lauki kendu dezakegu; *Editar* funtzioarekin, lauki bateko testua aldatu/zuzendu dezakegu. Gero, gainera, goiko botoietan, *Deshacer* daukagu; aukera horrekin, azken agindua desegin dezakegu, hanka sartu baldin badugu.

Lerrokatze lana amaitu dugunean, edo lanerako denbora gehiago ez badugu, lanaren emaitza gordetzeko modu bat baino gehiago dago. Saguarekin, *Guardar xml* erabil dezakegu, eta, hala, lerrotatu ditugun testu guztiak gordeko ditugu, banan-banan. Lanarekin aurrera jarraitu nahi dugunean, testu bakoitzaren xml artxiboa igo beharko dugu berriz, banaka-banaka. Bestalde, goialdeko funtzioetan, *Guardar tmx* dugu; aukera horrekin, lerrokatutako testu guztiak artxibo bakarrean gordeko ditugu (OLI programek erabiltzen duten formatu berbera da).



9. irudia: Lerrokatutako tmx artxiboa.

Hala, gordetako artxibo bakar hori kargatu egin dezakegu gero, lanarekin jarraitzeko. Hala ere, tmx formatoak badu arazo bat: ez du metadaturik gordetzen. Horregatik, Taligner 3.0. aldaeran, taf formatua gaineratu dugu, zeina *Guardar alineación* funtzioarekin aktibatzen den. Kasu horretan, testu guztiak artxibo bakarrean gordetzen ditugu, metadatuak barne. Gero, artxibo hori berriro kargatu daiteke, lanean jarraitzeko.

Lerrokatze lanak amaitu eta emaitza gorde badugu, datu basea elikatu behar dugu. Horretarako, testu bakoitza banan-banan igo behar dugu datu basera, saguaren funtzioen artean *Añadir a corpus* hautatuta. Aurreko prozedura hori hainbat aldiz errepikatu dezakegu corpusean sartu nahi ditugun testu bildumekin; hala, datu basea elikatzen jarraituko dugu.

Corpusa eratua dugunean, bilaketak egin ahal izango ditugu. Horretarako, goiko *Consultar corpus* atalean sartuko gara. Hor, bilaketa motorraren interfazea aurkituko dugu. Hala ere, bilaketak egin aurretik, corpusa modu egokian antolatu behar da, eta, horretarako, *Configurar corpus* atalean sartu, eta eragiketa batzuk egin beharko ditugu.



10. irudia: Corpusa konfiguratu. AleuskaFilo.

Hor, corpuseko testu bilduma osoa ikusi eta kudeatu ahal izango dugu. Bilaketa prozesuan arazorik ez izateko, testuen arteko harremanak zehaztu ahal izango ditugu, hau da, itzulpenen eta jatorrizkoen arteko harremanak, testuen gainean saguarekin *Editar metadatos y relaciones* aktibatuz. Bestalde, corpus osoan azpicorpusak zehazteko aukera ere badugu; hala, bilaketa-tan corpus guztia erabili beharrean, azpicorpus bakoitzean egin dezakegu bilaketa. Gainera, aurretik osatutako corpusak inportatu edo sortu duguna esportatzeko aukerak ere badauzkagu hor. Corpusa ongi eratu ondoren, bilaketak txukun egin ahal izango ditugu *Consultar corpus* funtzioa baliatuz.



11. irudia: Bilatzailea.

Bilaketa interfazea aurrez aurre izanda, argiago ikus daiteke metadatuaren funtzioa. Izan ere, testu bakoitza metadatuaz hornituta dagoenez, posible da, esate baterako, bilaketak egile jakin baten testuetan bakarrik egitea. Eta gauza bera egin daiteke hizkuntza, izenburu, itzultzaile, itzulpen modu eta generoarekin. Horretaz gain, hitza(k) bilatzeko orduan, aukera bat baino gehiago dugu: *Exacto*, *Contiene*, *Empieza con*, *Termina con*, alegia. Hala, hitza bere horretan, hitz zati bat, hitzaren hasiera edo hitzaren amaiera bilatu dezakegu. Gainera, hitz bateko bilaketa sinplea egiteaz gain, hitz konbinazioak bilatzeko aukera ere badugu (hiru hitzeraino), hitz horien arteko distantzia zehaztuz. Gainera, emaitzak bistartzeko unean, bilatu nahi dugun hitza edo esamoldea agertzen den esaldia bakarrik aukeratu dezakegu, edo aurreko eta ondorengo esaldi 1, 2 edo 3 ere ikus ditzakegu. Hona hemen adibide bat.



12. irudia: *Zunze* hitzaren bilaketa Canettiren *Die gerettete Zunge* testuan.

Bilatu d(ir)en hitza(k) horiz markatuta agertzen d(ir)a, eta, gainera, esaldia identifikatuta dago lehen zutabeen testuari eman zaion kodearekin eta esaldiaren zenbakiarekin. Gure kasuan, gainera, hitza agertzen den esaldiaren aurreko eta ondorengo esaldiak ere erakusten ditu bilaketak, testuinguru pixka bat gehiago izateko. Egiten diren bilaketak Excel fitxategi moduan gorde daitezke, gerora berriro kontsultatzeko.

Horretaz guztiaz gain, *Consultar ficheros* atala ere badugu. Hor, testuak analizatzeko aukerak nahi ditugu bildu, pixkanaka. Oraingoan, corpuseko testu bakoitzaren hustuketa terminologikoa egin daiteke, alfabetoaren arabera ordenatuta edo hitzen maiztasunaren arabera ordenatuta. Testu batean agertzen diren hitz guztiak eta euren maiztasuna ematen ditu horrek. Gainera, hitz kopuru gordina (*token*) eta hitz formen kopurua (*type*) kalkulatzen ditu.



13. irudia: Immanuel Kanten *Kritik der reinen Vernunft*.

Aurreko adibidean, 175.374 hitzez (*token*) osatutako testu bat dugu; tartean, 9.659 hitz ezberdin (*type*). Analisiaren emaitzak Excel taula batean gorde daitezke.

Amaitzeko, *Configurar Taligner* atala dugu. Hor, metadatueta sar daitezkeen informazioa egokitzeko tresna dugu, hala nola, hizkuntzak, itzulpen modua eta testuen generoa.

Etorkizunean

Etorkizunean programa garatzen jarraitzeko aukerarik balego, hauek lirateke egin beharreko hobekuntza batzuk: interfazea euskaratu eta ingeleseratu; bilatzailea testu lematizatuak erabiltzeko egokitu; testu analisirako tresneria zabaldu (Antconc programak eskaintzen dituen aukerak integratuz, besteak beste); eta abar.

BIBLIOGRAFIA

- Barambones, J. (2012). *Mapping the Dubbing Scene. Audiovisual Translation in Basque Television*. Bern: Peter Lang.
- Barambones Zubiria, J.; Manterola Agirrezabalaga, E.; Sanz Villar, Z.; Uribarri Zenekorta, I.; Zubillaga Gomez, N. (2015). «Itzulpen ikasketak eta euskara. Zenbait ekarpen Itzulpen Ikasketa Deskribatzailei, eta haratago», in *Traducimos desde el sur: actas del VI Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación*, José Jorge Amigo Extremera (arg.). Las Palmas kanaria Handikoa: Las Palmas kanaria Handikoko Unibertsitatea.
- Manterola, E. (2012). *Euskal literatura beste hizkuntza batzuetara itzulia. Bernardo Atxagaren lanen itzulpen moten arteko alderaketa*, Bilbo: Euskal Herriko Unibertsitatea.
- Sanz Villar, Z. (2015). *Unitate fraseologikoen itzulpena: alemana-euskara*, UPV/EHUko tesia.
- Toury, G. (1995). *Descriptive Translation Studies and beyond*, Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Zubillaga, N. (2009). «Alemanetik euskaratutako haur eta gazte literatura edo amaigabe istorioa?», *Senez*, 36: 231-241.
- (2012). «A corpus based descriptive study of German children's literature translated into Basque: Preliminary results», in *Translating Fictional Dialogue for Children and Young People*, Fischer, M. eta Naro W. (arg.), Berlin: Frank&Timme.
- ; Sanz, Z. eta Uribarri, I. (2015). «Building a trilingual parallel corpus to analyse literary translations from German into Basque», in *New directions in corpus-based translation studies*, Fantinuoli, C. eta Zanettin, F. (arg.), Berlin: Language Science Press.

Resumen

Taligner es un programa para la creación y análisis de corpus multilingües, desarrollado por un grupo de investigadores del Área de traducción de la EHU-UPV en colaboración con el informático Iñaki Albisua. La herramienta se ha utilizado ya en varios proyectos de investigación, lo que ha permitido, a su vez, desarrollar la aplicación. Taligner permite etiquetar y alinear textos narrativos y dramáticos, para poder analizar posteriormente los corpus creados, con ayuda de un buscador. Así mismo, permite gestionar diferentes corpus. En lo referente al análisis de los textos, por el momento la aplicación realiza el vaciado terminológico de los textos recogiendo la frecuencia de utilización de cada palabra.

Résumé

Le programme Taligner a été conçu par un groupe de chercheurs du département de traduction de l'Université du Pays Basque, avec l'aide de l'informaticien Iñaki Albisua, dans le but de créer et d'analyser des corpus multilingues. Il a été appliqué dans plusieurs projets de recherche et il a pu, en même temps, être développé grâce à ces projets. Il permet d'étiqueter et d'aligner les textes narratifs et dramatiques, pour analyser les corpus ainsi créés au moyen d'un chercheur. De plus, il permet également de gérer les différents corpus. Quant à l'analyse des textes, ce programme permet pour le moment de faire le dépouillement terminologique en comptabilisant la fréquence de chaque mot.

Abstract

The Taligner software was developed by a research team from the translation section at the University of the Basque Country in collaboration with IT expert Iñaki Albisua, with the aim of compiling and analyzing multilingual translation corpora. It has been used in several research projects and has been further developed thanks to those projects. It provides capabilities for tagging and aligning narrative and drama texts so that the compiled corpora may be analyzed using a search engine. Furthermore, the software makes it possible to manage different corpora simultaneously. With respect to text analysis, Taligner is currently able to produce term extractions of the texts, counting the frequency of each word.