

**PERPAUS ADBERBIALEN AGERPENA, MAIZTASUNA ETA KOKAPENA
EPEC-DEP CORPUSEAN**

Itziar Gonzalez-Dios, María Jesus Aranzabe, Arantza Díaz de Ilarraza

UPV/EHU / LSI / TR 02-2015

Perpaus adberbialen agerpena, maiztasuna eta kokapena EPEC-DEP corpusean

Itziar Gonzalez-Dios, María Jesús Aranzabe eta Arantza Díaz de Ilarraza

2015

Laburpena

Txosten honetan euskarazko perpaus adberbialen agerpenaren, maiztasunaren eta kokapenaren emaitzak aurkezten dira. Analisi hau egiteko, euskarazko EPEC-DEP zuhaitz-bankua edo Treebank-a eta Euskaltzaindiaren gramatika deskriptiboak erabili dira.

Laburpena

In this report we present the results obtained analysing the use, frequency of use and the position of adverbial clauses. This analysis has been performed in the Basque Dependency Treebank (BDT). We also have used the descriptive grammars of *Euskaltzaindia*, the Royal Academy of the Basque Language.

1 Sarrera

Txosten honetan euskarazko perpaus adberbialak ditugu aztergai. Zehazki, perpaus horien agerpena, maiztasuna eta kokapena aztertu ditugu. Horretarako, *Euskal Gramatika: Lehen Urratsak* (EGLU) gramatika deskriptiboan (Euskaltzaindia, 1999; Euskaltzaindia, 2005; Euskaltzaindia, 2011) agertzen diren egituren zerrendak atera ditugu eta egitura horiek euskarazko EPEC-DEP zuhaitz-bankuan edo Treebank-ean (Arantzabe, 2008) bilatu ditugu.

EPEC-DEP zuhaitz-bankuak Euskararen Prozesamendurako Erreferentzia Cor-puseko (EPEC) (Aduriz et al., 2006) 200.000 hitz hartzen ditu. Hitz horiek eskuz etiketatu dira Dependentsia Gramatikaren Teoria (Tesnière, 1959) jarraituz; etiketatze lan horretan esaldiko hitzak binaka lotuz esaldi bakoitzaren zuhaitz sintaktikoa (dependentsia-zuhaitza ere esaten zaio) lortu da. Zuhaitz horietan dependentsia-etiketen bidez adabegietan dauden hitzen arteko gobernatzaile/mendeko erlazioak irudikatu dira, eta bi hitzen arteko loturan mendekoak betetzen duen funtzio sintaktikoa adierazi da.

Txosten honetan bi esperimenturen emaitzak jaso dira: lehenengo esperimentuan, egitura bakoitza bere motaren barnean zenbat aldiz agertzen den (kopurua eta ehunekoa) aztertu da eta bigarren esperimentuan, egitura bakoitza zein kokapen den (aditza nagusiaren aurretik edo atzetik agertzen diren) analizatu da. Esperimentuetan kontaketak automatikoki egin dira, Perl programazio lengoia eta Unix komandoak erabilia. Emaitzak 2 atalean aurkezten ditugu. 3 atalean lan hau zertarako izan daitekeen erabilgarria eta etorkizuneko lanak aipatzen ditugu.

2 Emaitzak

Atal honetan esperimentuen emaitzak aurkezten ditugu. Taulak perpaus motaren (denbora, kausa...) eta aditz jokatuaren eta ez jokatuaren arabera daude antolatuta. Taulen lehenengo zutabearen, egitura zein den agertzen da; alegia, perpaus adberbialetako bakoitzari dagokion menderagailua edo erlazioa adierazi da. Egitura kurtsibaz idatzita egoteak esan nahi du, egitura hori ez dugula EGLUetan mota horren barruan aurkitu. Bigarren zutabearen egitura horren ingelesezko ordaina ematen da. Hirugarren zutabearen, egitura hori corpusean zenbat aldiz aurkitu dugun adierazten da eta, laugarrenean, egitura horren ehunekoa bere motako egiturak kontuan hartuta. Bosgarren eta seigarren zutabeetan, egitura hori aditz nagusiaren aurretik edo atzetik zenbatetan agertu den azaltzen da.

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoa	Aurretik	Atzetik
-enean	when	223	58,22	65,92	34,08
-ela	when/as	72	18,80	83,33	16,67
-elarik	when	1	0,26	100,00	0,00
-enetan	every time when	1	0,26	100,00	0,00
-en bakoitzean	every time when	1	0,26	100,00	0,00
-en guztietan	all the time when	0	0,00	-	-
-en aldikal	all the time when	0	0,00	-	-
-en aldiro	all the time when	0	0,00	-	-
-eneko	by the time	2	0,52	100,00	0,00
-enerako	by the time	4	1,04	75,00	25,00
-en orduko	by the time	0	0,00	-	-
-en bezain laster	as soon as	0	0,00	-	-
-en bezain sarri	as soon as	0	0,00	-	-
-en bezain agudo	as soon as	0	0,00	-	-
-en bezain fite	as soon as	0	0,00	-	-
-en ber	together with	0	0,00	-	-
-enaz batera	together with	0	0,00	-	-
-en baino lehen	before	0	0,00	-	-
-en aurrean	before	0	0,00	-	-
-en aitzinean	before	0	0,00	-	-
-en ondoan	after	0	0,00	-	-
-en ostean	after	3	0,78	66,67	33,33
-enetik	since	51	13,32	54,90	45,10
-enez gero	since	2	0,52	50,00	50,00
-en arte	until	2	0,52	0,00	100,00
-eno	during	0	0,00	-	-
-eino	during	0	0,00	-	-
-en bitartean	while	21	5,48	61,90	38,10
-en artean	while	0	0,00	-	-
-en arteko	while	0	0,00	-	-

1. taula: Denborazko perpaus jokatuak

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoak	Aurretik	Atzetik
-tzean	when	59	17,82	77,97	22,03
-tzerakoan	when	24	7,25	66,67	33,33
-tzeakoan	when	1	0,30	100,00	0,00
-tzearekin	when	1	0,30	0,00	100,00
-tzeari	when	0	0,00	-	-
-tzerat	when	0	0,00	-	-
-tu(k)eran	when	0	0,00	-	-
-tu aldiro	all the time when	0	0,00	-	-
-tu bakoitzean	every time when	0	0,00	-	-
-tu guztian	all the time when	0	0,00	-	-
-tu ahala	as soon as	10	3,02	100,00	0,00
-tu arau	as soon as	0	0,00	-	-
-tzerako	by the time	4	1,20	100,00	0,00
-tu orduko	by the time	9	2,71	88,89	11,11
-tu bezain laster	as soon as	1	0,30	100,00	0,00
-tu bezain pronto	as soon as	2	0,60	100,00	0,00
-tu eta berehala	as soon as	1	0,30	100,00	0,00
-tu eta laster	as soon as	0	0,00	-	-
-tuaz	together with	0	0,00	-	-
-tzearekin bat(era)	together with	0	0,00	-	-
-tu berri(t)an	just when	0	0,00	-	-
-tu baino lehen	as soon as	4	1,20	100,00	0,00
-tu aurretik	before	39	11,78	71,79	28,21
-tu aitzinean	before	0	0,00	-	-
-tu gabe	without	0	0,00	-	-
-tu eta	after	6	1,81	100,00	0,00
-tu eta gero	after	25	7,55	52,00	48,00
-tuta	once	4	1,21	75,00	25,00
-tutakoan	once	8	2,42	87,50	12,50
-tu ondoan	after	4	1,21	50,00	50,00
-tu ostean	after	83	25,07	80,72	19,28
-tuz gero	after	1	0,30	100,00	0,00
-turik	past participle	0	0,00	-	-
-tu arte	until	23	6,94	43,48	56,52
-tu artean	until	4	1,21	100,00	0,00
-tu bitartean	while	4	1,21	50,00	50,00
-tzeraino	until	1	0,30	0,00	100,00
-tu aitzin	before	4	1,21	75,00	25,00
-tu osteko	after	1	0,30	100,00	0,00
-tu bezperan	on the eve of	1	0,30	0,00	100,00
-tu ondotik	after	6	1,81	66,67	33,33
-tzeaz	to be about to	1	0,30	100,00	0,00

2. taula: Denborazko perpaus ez jokatuak

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoa	Aurretik	Atzetik
-elako((t)z)	because	190	26,91	11,05	88,95
-elakoan	supossing	0	0,00	-	-
bait-	since	282	39,94	6,38	93,62
zeren eta...(bait-/-(e)n)	since	4	0,57	0,00	100,00
... eta	since	51	7,22	74,51	25,49
zeren	for	1	0,14	0,00	100,00
zergatik	why	0	0,00	-	-
ezen (...) (bait-)	since	1	0,14	0,00	100,00
-enez gero(ztik)	since	3	0,42	66,67	33,33
-enez	as	169	23,94	77,51	22,49
-en legez	as	0	0,00	-	-
-ela kausa	to be the cause of	1	0,14	0,00	100,00
-ela bide	to be the cause of	0	0,00	-	-
-ela medio	to be the cause of	4	0,57	75,00	25,00

3. taula: Kausazko perpaus jokatuak

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoa	Aurretik	Atzetik
-tzeagatik	because	2	9,52	50,00	50,00
-tzearren	because	2	9,52	50,00	50,00
-turik	past participle	0	0,00	-	-
-tutakoan	when	0	0,00	-	-
-tuz	-ing	3	14,29	66,67	33,33
-tuta	past participle	2	9,52	50,00	50,00
-tuz gero	since	11	52,38	54,55	45,45
-tzearekin	with	1	4,76	100,00	0,00

4. taula: Kausazko perpaus ez jokatuak

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoa	Aurretik	Atzetik
nahiz (eta) -(en/-ela/ba-)	although	15	11,72	40,00	60,00
-en arren	though	56	43,75	76,79	23,21
ba- (...) ere	though	57	44,53	85,96	14,04

5. taula: Kontzesiozko perpaus jokatuak

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoa	Aurretik	Atzetik
nahiz eta... /-tu/ø	although	28	23,93	39,29	60,71
-tu arren	though	77	65,81	76,62	23,38
-tugatik	though	1	0,85	100,00	0,00
-tuta (gabe/ezta) ere	without ... too	2	1,71	50,00	50,00
-turik (gabe/ezta) ere	without ... too	0	0,00	-	-
-tuz gero ere	after ... too	0	0,00	-	-
-tzearren	although	5	4,27	40,00	60,00
-tuz gero	although	2	1,71	100,00	0,00
-ik ere	although	2	1,71	100,00	0,00

6. taula: Kontzesiozko perpaus ez jokatuak

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoa	Aurretik	Atzetik
-ela	as	45	18,07	73,33	26,67
-elarik	as	47	18,88	23,40	76,60
-en moduan/ra	the way	1	0,40	100,00	0,00
-en arabera(n)	the way	0	0,00	-	-
-en eran/ra	the way	0	0,00	-	-
-en antzera	similar to	1	0,40	0,00	100,00
-en moldean/ra	the way	0	0,00	-	-
-en gisan/ra	the way	0	0,00	-	-
-en bezala	like	29	11,65	68,97	31,03
-en legez	like	2	0,80	0,00	100,00
-enez	as	124	49,80	76,61	23,39

7. taula: Moduzko perpaus jokatuak

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoa	Aurretik	Atzetik
-tuz	-ing	401	33,31	48,13	51,87
-tuta	past participle	501	41,61	61,68	38,32
-turik	past participle	86	7,14	63,95	36,05
-tu gabe	without	59	4,90	38,98	61,02
-tu barik	without	2	0,17	50,00	50,00
-tu ezta	without	0	0,00	-	-
-tu ordez	instead of	5	0,42	80,00	20,00
-tu ordean	instead of	0	0,00	-	-
-tu aginean	instead of	0	0,00	-	-
-tu aginik	instead of	0	0,00	-	-
-tu ahala	as soon as	5	0,42	80,00	20,00
-tu arau	as soon as	0	0,00	-	-
-tu beharrean	willing	7	0,58	85,71	14,29
-tu nahirik	willing	1	0,08	0,00	100,00
-tu nahian	willing	8	0,66	25,00	75,00
-tu gurarik	willing	0	0,00	-	-
-tu ezinik	in a impossible way	1	0,08	0,00	100,00
-tu ezinda	in a impossible way	3	0,25	33,33	66,67
-tu ezinean	in a impossible way	1	0,08	100,00	0,00
-tu beharrez	needing	0	0,00	-	-
-tzeko zorian	to be about	3	0,25	33,33	66,67
-tu hurran	to be about	0	0,00	-	-
-tzeko moduan	the way	8	0,66	87,50	12,50
-tzeko gisan	the way	0	0,00	-	-
-tzeko eran	the way	1	0,08	100,00	0,00
-tzeko maneran	the way	0	0,00	-	-
-tzeko moldean	the way	0	0,00	-	-
-tzeke	without	2	0,17	50,00	50,00
-tzekotan	in case of	8	0,66	87,50	12,50
-tu bezala	the way	20	1,66	65,00	35,00
-tzera	with the aim	41	3,41	92,68	7,32
-tzeaz	-ing	16	1,33	87,50	12,50
-tzeaz gain	apart from	19	1,58	100,00	0,00
-tik	from	6	0,50	100,00	0,00

8. taula: Moduzko perpaus ez jokatuak

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoa	Aurretik	Atzetik
subjuntiboa	in order that	56	100,00	17,86	82,14
-tzeko(tz)	(in order) to	791	88,38	40,20	59,80
-tzekotzat	(in order) to	0	0,00	-	-
-tzearren	(in order) to	15	1,68	46,67	53,33
-tzeagatik	(in order) to	0	0,00	-	-
-tze alde(ra)	(in order) to	0	0,00	-	-
-tzekotan	(in order) to	0	0,00	-	-
-tzeko asmotan	with the intention of	1	0,11	100,00	0,00
-tzeko intentziotan	with the intention of	0	0,00	-	-
-tzeko intentzioarekin	with the intention of	0	0,00	-	-
-tzera	(in order) to	88	9,83	61,36	38,64

9. taula: Helburuzko perpaus jokatu eta ez jokatuak

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoa	Aurretik	Atzetik
mailakatzaila (non/ezen) ... bait-	so that	7	58,33	14,29	85,71
mailakatzaila (non/ezen) ... -en	so that	5	41,67	0,00	100,00

10. taula: Ondoriozko perpaus jokatuak

Euskarazko egitura	Ingeleseko ordaina	Kopurua	Ehunekoa	Aurretik	Atzetik
ba-	if	249	100,00	81,12	18,88
non (eta) ez ... -en	if	0	0,00	-	-
-tuz gero((z)tik)	if	34	70,83	85,29	14,71
-tu ezker(an/k/tino/((z)tik))	if	1	2,08	0,00	100,00
-tu ezik	unless	0	0,00	-	-
-tu ezean	unless	2	4,17	50,00	50,00
-tzekotan	in the case of	8	16,67	100,00	0,00
-tzekoz	in the case of	0	0,00	-	-
-tzez gero	in the case of	0	0,00	-	-
-tuenean	if	0	0,00	-	-
-tzera(t)	if	3	6,25	100,00	0,00

11. taula: Baldintzazko perpaus jokatu eta ez jokatuak

3 Ondorioak

Txosten honetan euskarazko perpaus adberbialen agerpenaren, maiztasunaren eta kokapenaren azterketa egin dugu. Lan hau, lan linguistiko deskriptiboa izateaz gain, Hizkuntzaren Prozesamenduko aplikazio aurreratueta baliagarria da. Adibidez, testuen sinplifikazioan (Aranzabe, Díaz de Ilarraza, and Gonzalez-Dios, 2012; Gonzalez-Dios, 2014) zein egitura diren ezezagunak adierazten digu, zeintzuk sinplifikatu behar diren jakiteko; esaldien hurrenkera zehazteko ere probetxugarria da. Hizkuntzaren sorkuntzan zein egitura erabili behar diren aukeratzeko ere balio du.

Etorkizunean, esperimentu horiek corpus handiagoetan egitea komeniko litza-teke. Honela, hemen aurkitu diren joerak baieztatuko genituzke eta aurkitu ez ditugun egituren berri izan genezake.

References

- Aduriz, Itziar, María Jesús Aranzabe, Jose Mari Arriola, Aitziber Atutxa, Arantza Díaz de Ilarraza, Nerea Ezeiza, Koldo Gojenola, Maite Oronoz, Aitor Soroa, and Ruben Urizar, 2006. *Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for automatic processing*, volume 56, pages 1–15. Rodopi.
- Aranzabe, María Jesús. 2008. *Dependentzia-ereduan oinarritutako baliabide sintaktikoak: zuhaitz-bankua eta gramatika konputazionala*. Ph.D. thesis, Euskal Filologia Saila (UPV/EHU).
- Aranzabe, María Jesús, Arantza Díaz de Ilarraza, and Itziar Gonzalez-Dios. 2012. First Approach to Automatic Text Simplification in Basque. In Luz Rello and Horacio Saggion, editors, *Proceedings of the Natural Language Processing for Improving Textual Accessibility (NLP4ITA) workshop (LREC 2012)*, pages 1–8.
- Euskaltzaindia. 1999. Euskal Gramatika Lehen Urratsak. In V, (*Mendeko perpausak-1, osagarriak, erlatiboak, konparaziozkoak, ondoriozkoak*). Euskaltzaindia, Bilbo.
- Euskaltzaindia. 2005. Euskal Gramatika Lehen Urratsak. In VI, (*Mendeko perpausak-2, baldintzazkoak, denborazkoak, helburuzkoak, kausazkoak, kontseziozkoak eta moduzkoak*). Euskaltzaindia, Bilbo.

- Euskaltzaindia. 2011. Euskal Gramatika Lehen Urratsak. In *VII, (Perpaus jokatugabeak: denborazkoak, kausazkoak eta helburuzkoak, baldintzazkoak, kontzesiozkoak, moduzkoak, erlatiboak eta osagarriak)*. Euskaltzaindia, Bilbo.
- Gonzalez-Dios, Itziar. 2014. Euskarazko testuak errazten: euskal testuen simplifikazio automatikoa. In Itziar Aduriz and Ruben Urizar, editors, *Euskal Hizkuntzalaritzan egungo zenbait ikerlerro. Hizkuntzalari euskaldunen I. to-paketa*. Udako Euskal Unibertsitatea.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Librairie C. Klincksieck, Paris.