



ELAN analisi multimodalerako lanabesa

Jasone Salaberria Fuldain

► **To cite this version:**

Jasone Salaberria Fuldain. ELAN analisi multimodalerako lanabesa. I. Jardunaldiak Ondareaz eta Hezkuntzaz, Nov 2012, Euskal Herriko Unibertsitatea - Leioa, Espainia. <artxibo-00757009>

HAL Id: artxibo-00757009

<https://artxiker.ccsd.cnrs.fr/artxibo-00757009>

Submitted on 25 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ELAN analisi multimodalerako lanabesa

Jasone Salaberria Fuldain
IKER UMR 5478
j.salaberria@wanadoo.fr

Laburpena

ELAN, EUDICO Linguistic ANnotator-en akronimoa da, eta EUDICO, Nijmegen-go –Herbeherak- Max Planck Psikolinguistika Institutuaren proiektu baten izena da (*EUropean DIstributed CO*rpora), zeinen egileak Hennie Brugman eta Albert Russel diren.

Programa hau analisi multimodalerako garatua izan da, ikus-entzunezko baliabideetatik etiketatze konplexuak egiteko. Eduki idatzia Unicode moldean dago eta transkripzioa XML formatuan atxikitzen du. Bestalde, ELANek dagoen software-n audio edo/eta bideo fitxategi-irakurleak erabiltzen ditu, Windows Media Player, QuickTime ou JMF- (Java Madia Framework) bezelakoak. Ondorioz, audio eta bideo formatu barietate zabala jasan dezake. Java programazio hizkuntzan idatzia dago eta bere baliabideak erabilera ez-komertzialetarako eskuragarri dira.

Software honek beste formatu batzuetara inportazio/esportazioa egiteko aukera ematen du.

Hitz gakoak : analisi multimodal, corpus, etiketatze.

Resumen

ELAN, es el acrónimo de EUDICO Linguistic ANnotator. EUDICO –*EUropean DIstributed CO*rpora- es un proyecto del Instituto de Psicolingüística Max Planck de Nijmegen -Holanda-. Siendo sus autores Hennie Brugman y Albert Russel.

Esta herramienta ha sido creada para análisis multimodales, para etiquetados complejos a partir de recursos audio-visuales. El contenido escrito lo guarda en Unicode y la transcripción en formato XML. Además, ELAN utiliza los lectores de archivo audio y/o video de los softwares existentes, como Windows Media Player, QuickTime ou JMF (Java Media Framework). En consecuencia, ELAN soporta una extensa variedad de formatos audio y video. Este programa está escrito en el lenguaje de programación Java y sus recursos están disponibles para utilizaciones no comerciales.

Este programa posibilita la importación/exportación a otros formatos.

Palabras claves: análisis multimodal, corpus, etiquetado

1. *Analisi multimodala*

Solasaldi-analisi multimodala Charles Goodwin-en¹ inspiraziopean garatu da 1979 eta 2000 urteen artean. Ahozko hizkuntzalaritza, hala nola elkarrekintzazko edo interakziozko hizkuntzalaritza, teknologien garapenarekin batera garatu dira. Zeinei esker, lehenik audioen grabaketa eta ondotik irudiena, ahal izan den hizkuntzaren eta ekintza sozialen erabilera testuinguruan grabatzea. Halere, eskeintza teknologiko hori bera ez da izan aski elkarrekintzazko ahozko hizkuntzalaritza sortzeko.

Gaur egun gero eta hizkuntza ikertzaile gehiagok -baita sozial eta giza zientziakoek ere- erabiltzen du bideo grabaketa hizkuntza komunikazioa testuinguru arruntean analizatzeko. Baina grabatutako datuak, sarritan, inplizitu eta guti kontzeptualizatuak gelditzen dira, eta helburu desberdinetarako erabiliak izaten dira:

- deiktikoen erreferentziez edo egoeraren inplizituez jabetzeko;
- objektu bat tratatzeko, adibidez, jestu ko-berbalak analizatzeko;
- hizkuntzaren ikusmolde desberdin batean oinarrituriko analisi linguistikoak egiteko, non hizkuntza, linguistikoa baino zabalagoa kontsideratzen den, non hizkuntza eta jestualitatea ezin diren bereizi, artifizialki ez bada, eta non antolaketa funtsez multimodala den.

ELAN hirugarren ikusmolderako asmatua izan da. Non kontsideratzen diren lengoiaia eta jestuak bereizezinak. Baliabide multimodal desberdinen adierazgarritasuna 70etako Kendon eta Goodwinen lanen ondotik hasi zen sistematikoki aztertzen. Baliabide multimodalak hauexek izanik:

- baliabide linguistikoak;
- aho eta soinu baliabideak (irri, aho-zaratak...);
- jestu baliabideak (eskuaren mugimenduari lotuak);
- baliabide bisualak (begirada mugimenduari lotuak);
- gorputz baliabideak (gorputzaren posturei lotuak);
- aurpegi baliabideak (mimikari lotuak);
- gauzen manipulazioa (objektu, dokumentu, lanabesak...).

Eta baliabide horien ezaugarriak dira, alde batetik, partehartzailearen arabera definituak direla, bere orientazioa elkarrekintzan (hurbilketa emikoa) emanak; bestalde, interakzioan hartzen dutela bere adierazgarritasuna eta zentzua. Gainera, ekintzaren denboran eta hizketaren interakzioan estuki ainguratuak dira. Eta testuinguruari lotuak dira.

Beraz, ez bakarrik hizketa, baizik eta eskuragarri eta ikusgarria den informazioak ere interesa du, multimodalitatea baita elkarrekintzan, sinergian diren kanalak, formak, adierazpen-funtzioak elkartzea, zentzuaren eraikuntzara ongi heltzeko².

¹Ondoko artikuluen izenburuek ongi erakusten dute Goodwin-en multimodalerako grina: "Sententziaren eraikuntza interaktiboa solasaldi naturalean" (1979); "Elkarrizketaren antolaketa. Hiztunen eta entzuleen arteko elkarrekintza" (1981); "Jestua eta ko-partaidetza hitzen bilaketa jardueran (1986); edo Jarduera eta irudia giza interakzioan" (2000).

² Dominique Boutet In Atelier DEGELS 2012: Défi Geste Langue des Signes, 36 or.

2. ELANen ezaugarri teknikoak

Erabiltzaileak nahi beste iruzkin gehi ditzaiete irudia eta soinu baliabideei. Iruzkina : perpaus, hitz, glosa, itzulpen edo edonolako deskribaketa bat.

Lerroetako eduki idatzia *Unicode*³ moldea erabiltzen du, eta transkripzioa XML⁴ formatuan atxikitzen du. Izatez, XML funtzionalitate aurreratuenetatik arindua izan den SGML⁵ lengoaiaren ekoizpena da. Hura, ikasteko eta egunero erabiltzeko konplexua izanik, HTML bertsio arinago bat garatu zen, baina azken hau, hainbat egokitzapen eginagatik ere, ezin izan zen hedatu mugarik gabe garatzaileen premiak emendatu arau. Orduan zen XML sortu.

XML, bereziki, datuen eta dokumentuen trukaketarako egokitua da, aplikazio eremu guztietan erabilgarria baita. Informazio trukaketarako formatu amankomun bat edukitzea abantaila handia da, eta praktikan, XML ardurako etiketatze hizkuntza bezala inposatzen ari da.

Programa honek audio eta bideo formatu barietate zabala jasan dezake : *.wav* audio fitxategiak, *.mpg1*, *.mpg2*, *.mov*, *.avi*, *.wmv* bideo fitxategiak. Horrela, adibidez, *.mts* luzapena duten bideo fitxategiak MTS bihurtuz *.wav* fitxategiak sortu eta bideoa fitxategia *mpeg* codec batez trinkotzen ahal da *.mpg1* edo *.mpg2* luzapeneko fitxategian bilakaturik.

Lau bideo-artxibo arte batera dakizkioke iruzkin dokumentu bakarrari. Eta erabiltzaileak nahi adina iruzkin gehi ditzaiete ikus-entzunezko baliabideei.

Bestalde, ELANek hainbat formatutara inportazio eta esportazioa egiteko aukera ematen du. Beste programa batzuk sorturiko formatuak ELANera inporta daitezke. Halere, inportazio motaren arabera parametro desberdinak berrikusi behar dira. Batzuetan emaitzak ez dira guztiz fidelak jatorrizko fitxategiarekin. Beraz, transkripzioaren lerrokatzea ELANen pean dagoelarik berrikustea komeni da.

Eta beste programa batzuekin baliatzeko, ELANekin lerrokatutako transkripzioak hainbat formatutara esporta daitezke.

3. Programaren instalatzea

<http://www.lat-mpi.eu/tools/elan/download> helbide elektronikoan deskarga daiteke ELAN programa. Bertan atzematen da eskuliburua ere linean edo PDF formatuan.

Windows pean instalatzea, erraza da, aski da exekutagailua abiaraztea eta aginduak betetzea.

³ *Unicode*, mundu mailan hizkuntza desberdinetako testu trukaketa haizutzen duen informatika estandar bat da.

⁴ *Extensible Markup Language*, “markatzeko lengoia hedagarri”

⁵ *Standard Generalized Markup Language*, “markatzeko lengoia orokor estandar”

Mac pean, aldiz, arazo batzuk eman ditzake, ordenagailuan aitzinetik instalatua den *Javaren* bertsioaren arabera. Ordenagailuaren *Java* bertsio berdina duen ELAN bertsioa instalatzea aholkatzen da.

4. Dokumentuen kudeaketa

Ikus-entzunezko fitxategiak ELANekin erabiltzeko, lehenik, audio edo/ta bideo fitxategiak gordeak izan behar dira dosier batean, hau da, .wav luzepenezko audio artxiboak eta .pmg1 edo .mpg2 edo .mov edo .avi edo .wmv luzepena duten bideo fitxategiak.

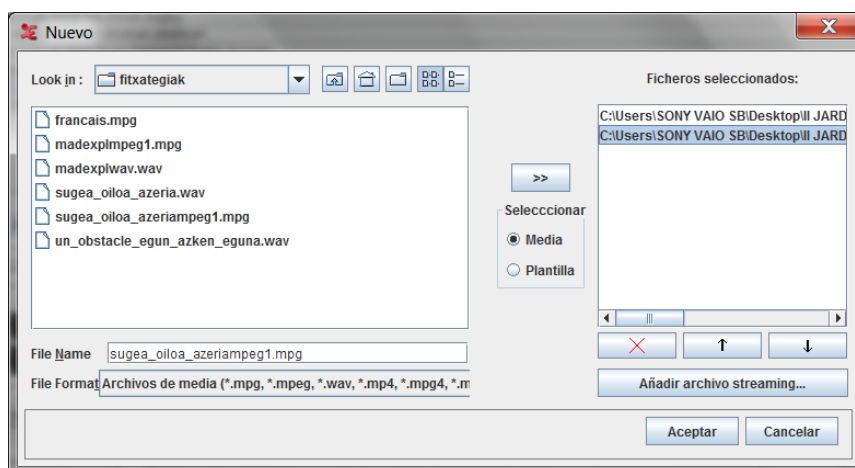
4.1. ELANen abiaraztea

ELAN ikonoan klik eginez, hastapeneko leihoa agertzen da. Non fitxategi menuan hainbat eskaintza aurkezten duen, bestek beste dokumentu *berria* sortzea.

4.2. ELAN fitxategi berri baten sortzea

Lanerako baliatuko diren audio/bideo fitxategiak dosier berdinean sailkatuak izatea komeni da. Horrela, *ireki* edo *berria* eskatzen delarik, ELANek dosier horiek zabalduko ditu zuzenean.

- a) *Fitxategi* menutik --> *Berria* aukeratu iruzkin berri bat abiarazteko. ELANek lehenik, eginen diren iruzkinei dagozkien bideo/audio fitxategien kargatzea eskatuko du leiho hau eskainiz:



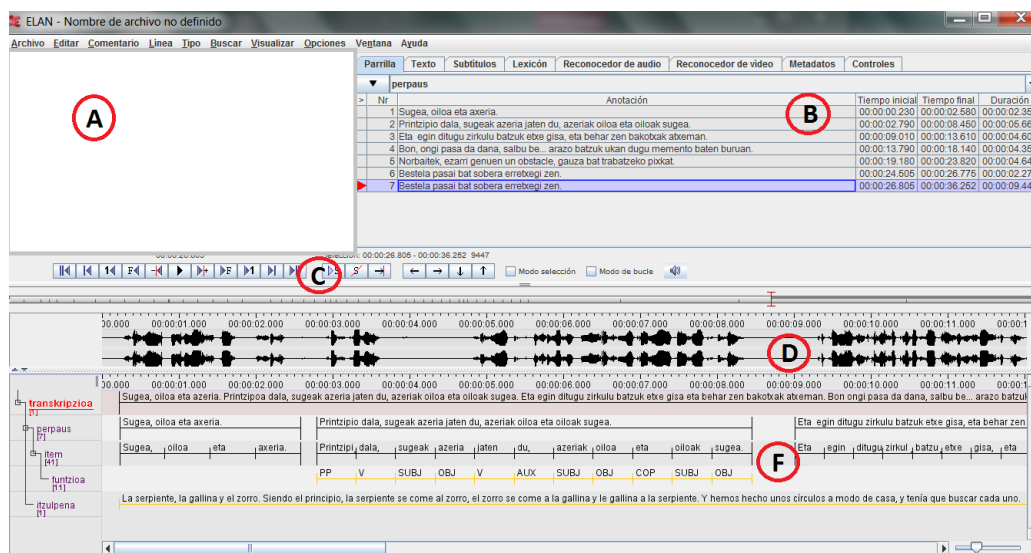
1 Irudia: Fitxategien hautaketarako elkarriketa-kuadroa

- b) *Fitxategi mota* : *Media fitxategiak* (.mpg, .mov, .wav).
c) Elkarriketa-kuadroaren ezkerrean hautatu behar dira bideo eta/edo audio fitxategiak eta eskuin aldera eraman [>>] botoia erabiliz, edo fitxategiaren gainean bi klik eginez.
d) *OK* balidatu.
e) Iruzkin lanean hasteko prest dago tresna. Gordetzeko *Fitxategi* menutik *gorde-n* klik egin, eta .eaf luzapena duen fitxategia sortuko du.

Hurrengo baten lanean jarraitzeko, aski da *Fitxategi* menuan *Ireki* sakatzea elkarriketa kuadroan .eaf luzepeneko artxiboak eskaintzeko. Oraingoan bideo eta soinu artxiboak batera irekiko dira.

4.3. Interfazearen aurkezpena

ELANen interfazeak, Imaia 2k erakusten duen bezala, bost eremu eskeintzen ditu:



2 Irudia: ELANen lanerako interfazea

- bideo irakurgailuaren eremua;
- iruzkinak molde desberdinetan afixatzeko eremua: Taula (Parrilla, Grille), Testua... eta Kontrolak erlaitzekin;
- bideo eta/edo audio kontrolagailuaren eremua;
- audio seinalearen eremua: audio artxibo bat inportatzen bada bere bideo artxiboarekin, seinalearen oszilograma ikus daiteke. Seinalearen gainean eskuineko botioarekin klik eginez, laster-menua agertarazten da eta seinalearen zoom maila bertikala edo horizontala modifika daiteke;
- iruzkinetarako lan-partizioaren eremua.

Iruzkina diogularik, perpau, hitz, glosa, transkripzio fonetiko, itzulpen edo edonolako deskribaketa esan nahi dugu. Hauek geruza ainitzetan sortuak dira eta ikus-entzunezko baliabideei denboran lerrotatuak. ELANek iruzkinen ikuspegi desberdinak eskeintzen ditu, eta ikuspegi bakoitza dagokion iturburuari lotua eta sinkronizatua dago.

5. Etiketatzeara

Etiketatzeara hiru alditan egiten da : a) *mota linguistikoen eta iruzkin lerroen* definizioa, b) *denbora segmentuen* hautapena, d) *iruzkinen* idaztea.

Transkripzio soila egin nahi bada, aski da lerro bat sortzea eta idaztea besterik gabe. Baina proiektuak eremu deberdinetako analisiak eskatuko baditu, proiektua ongi definitua

izan behar da aitzinetik -helburua betetzeko zenbat eremu analizatu beharko diren, zein den eremuen arteko harremana eta hierarkia, etiketatze batzuk egiteko hiztegi mugatu bat erabil daitekeen, etab-.

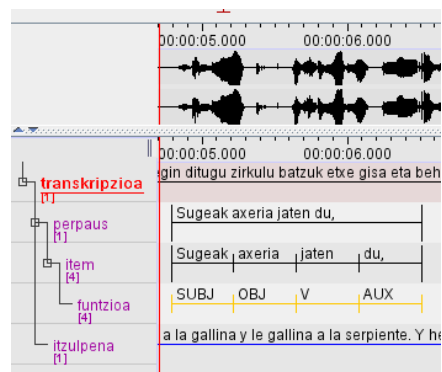
Jakin beharra dago lerroak bi motakoak izan daitezkeela:

- *Independienteak*, barnean dituzten informazioak denbora tarteari lotuak zuzenean direnak, hau da, denbora ardatzarekin sinkronizaturik (time-alignable).
- *Erreferentzialak*, barnean dituzten informazioak beste *Lerro* bateko (*Lerro Nagusia*, Parent, Línea Madre deitua) iruzkinei lotuak direnak. Ez dira zuzenean denbora ardatzari lotuak. Haietariko batzuk egon daitezke hala ere, baina tira nagusiaren denbora tarte zehatzaren barnean bakarrik.

Iruzkinak agertzeko eremuan, *lerro* erreferentzialaren etiketak *lerro nagusiaren* kolore berdina izanen du.

Hierarkia kateatuak sor daitezke, adibidez, A *lerroa* B *lerroaren lerro nagusia* izan daiteke, eta B *lerroa* C *lerroaren lerro nagusia*, etabar, Imaia 5ean ikus daitekeen legez:

- *transkripzioa* lerroa independentea da,
- *perpaus* lerroa *transkripzioa* lerroaren menpe da,
- *item* lerroa *perpaus* lerroaren menpe da,
- *funtzioa* lerroa *item* lerroaren menpe da,
- *itzulpena* lerroa *transkripzioa* lerroaren menpe da.



3 Irudia: Lerroen arteko hierarkia

Lerro nagusien eta menpekoen lotura hain da handia, non nagusiei egin aldaketek menpekoan eragingo duten, baina ez alderantziz.

Lerro nagusi bat ezabatzen bada, menpeko guztiak ere ezabatzen dira. Gauza bera, lerro nagusi baten iruzkina kentzen bada, haren *menpeko lerroaren* iruzkina ere desagertzen da. *Lerro nagusi* baten denbora-tartea aldatzen bada, haren *menpeko lerro* guztien denbora-tartea ere aldatuko da.

5.1 Mota linguistikoaren definizioa: hiztegi kontrolatua + estereotipoa

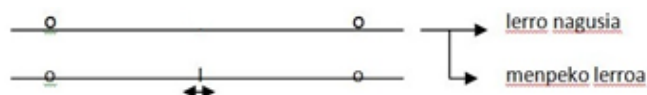


5.1.1. Estereotipoak

Lerro bakoitzarentzat *Mota (Tipo, Type) linguistiko* bat finkatzen da. *Mota linguistiko* bakoitzak, mota hortako *lerro* guztiei aplikatzen zaizkien betebeharrak batzuk zehazten ditu. Betebeharrak horiek inplizitoki laburtuak dira *estereotipoetan*.

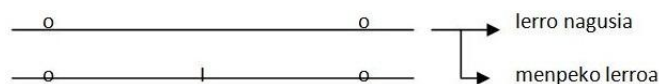
Mota linguistikoek bost estereotipo eskeintzen dituzte:

- *Bat ere ez (None)*: Iruzkina zuzenean denbora ardatzari lotua da, hau da, iruzkina lerro independente baten sartzen da.
- *Denbora azpizatiketa (Time Subdivision)*: Lerro nagusiaren iruzkin bat menpeko lerroan azpiunitatetan zati daiteke, azpiunitateak zeinek, bere aldikoz, denborari lotuak izan daitezkeen.



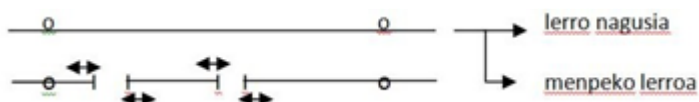
Adibidez, lerro nagusiko perpausak sintagmetan zati daitezke menpeko lerroan, bakoitza dagokion denbora tarteari lotua izanik. Barne zatiketa bakoitza arrasta daiteke. Horrelako lerrotako iruzkinak denborazko sinkronizazioa dute.

- *Azpizatiketa sinbolikoa (Symbolic Subdivision)*: *Denbora azpizatiketaren* antzekoa, baina honetan unitatepekoak ez dira denboran sinkronizatuak. Adibidez, lerro nagusiaren hitz bat morfema indibidualetan zatitua izan daiteke (ez dutenak denborazko sinkronizaziorik ukanen).



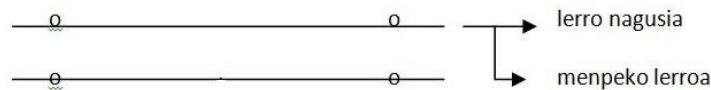
Estereotipo honetan barne zatiketak ezin dira mugitu.

- *Barnean (Included In)*: menpeko lerroaren iruzkin guztiak lerro nagusiaren iruzkinen mugen barnean erortzen dira. Baina hutsak egon daitezke menpeko iruzkinen artean. Adibidez, isilguneak dituen perpaus baten iruzkina, menpeko lerroan hitzetan moztu daiteke isilguneei espazio hutsak dagozkiela.



Barnean estereotipoan barne zatiketak mugitu daitezke.

- *Asoziazio Sinbolikoa (Symbolic Association)*: menpeko lerroaren iruzkinak ezin du lerro nagusiaren iruzkina zatikatu. Estereotipo hau erabiltzen da, adibidez, lerro nagusiko perpausaren itzulpen laxoa egiteko.



5.1.2. Hiztegi kontrolatua

Edizio menuan --> *Hiztegi kontrolatua editatu* (Editar vocabularios controlados, Editer le vocabulaire contrôlé, Edit controlled vocabularies) botoiari klik egin (4 Irudia).

Izen hori hartzen du errepikatzen diren etiketekin aitzinetik osatzen den hiztegiak.

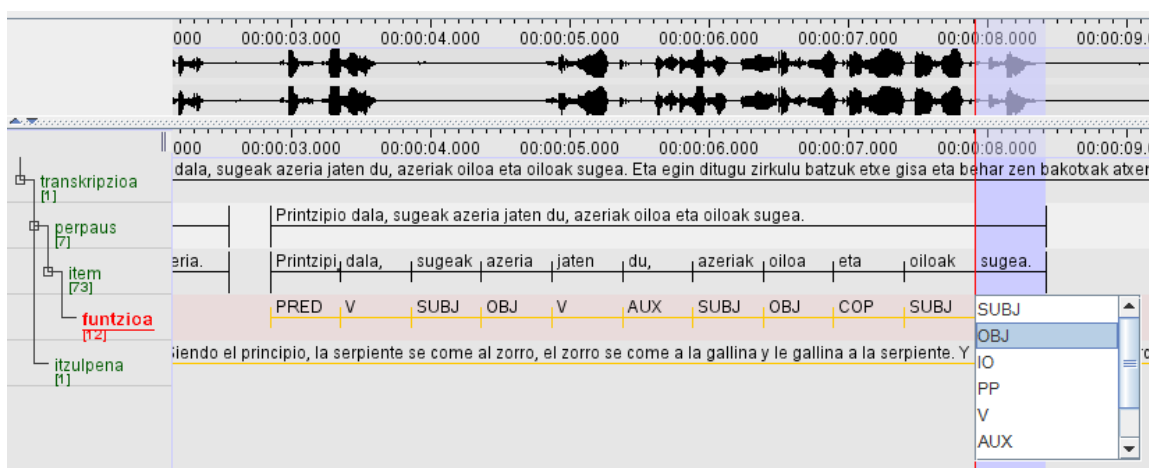
Adibidez, proiekturako funtzio gramatikalak aztertzea garrantzizkoa bada, *funtzioa* lerro bat sotzen da.

Funtzio gramatikalen zerrenda itxia denez, *Funtzio gramatikalak* izeneko hiztegi kontrolatua sor daiteke aitzinetik. Horrela *funtzioa* lerroan iruzkinak sartzeko unean, hedatzen den leiho bat ageriko da funtzio gramatikalak eskainiz. Aski da *itemari* dagokion funtzioa aukeratzea (5 Irudia).

Bide hau lagungarri izan daiteke gauza bertsuak idatzi behar direnean baita akatsak saihesteko ere.



4 irudia: Hiztegi kontrolatua sortzeko elkarrizketa-kuadroa



5 Irudia: Funtzioa lerroko iruzkinak sartzeko hiztegi kontrolatuaren eskeintzak

5.2. Lerroen definitzea

Mota linguistikoak sortuak direlarik lerroak defini daitezke. *Lerro* bat sortzeko *Lerro* (*Línea*, *Acteur*, *Tier*) menuan zapatu eta *Gehitu lerro berri bat* (ctrl+T) (*Añadir Línea Nueva*, *Add New Tier*) balidatu, elkarrizketa-kuadroko ondorengo eremuak agertuko dira:

- *Lerroaren izena*: lerroaren buruan agertuko den izena, etiketatze eremuaren izena.
- *Parte-hartzailea*: hiztunari buruzko informazioa (ez da ELAN uneko leihoan ageriko).
- *Iruzkinegilea*: interesgarria etiketatzaile desberdinak parte hartzen dutenean.
- *Lerroaren Lerro Nagusia (Línea Madre, Parent Tier)*: dagokion lerro nagusia aukeratu.
- *Mota linguistikoa*: dagokion mota aukeratu (lerroa eta mota linguistikoa izen bera).
- *Lehenetsi hizkuntza*: ez dakit zer eragin ekar dezakeen eremu honek, euskara ez baitago aukeran, baina badirudi ez duela eragin handirik.
- *Gehitu (Añadir, Add)* balidatu klik eginez.
- *Opzio gehiago* funtzioak: letra-tipo eta koloreak aldatzeko balio du.

Lerro baten gehitzean, elkarrizketa-kuadroa ez da automatikoki hestean; beste lerro batzuk ere gehi daitezke. Lan leihora itzultzeko, *Itxi* balidatu behar da.

Nahi den beste lerro defini daiteke, eta gorago esan bezala, aholkatzen da hastapenetik ongi pentsatzea zer informazio mota aztertu behar den nahi den helburua lortzeko, hau da, lerro bakoitzarentzat zer iruzkin sartuko den eta lerro bat baino gehiago badira, lerroen arteko hierarkia ongi zehaztua izatea. Hala ere, geroago ere *menpeko lerro* baten *lerro nagusia* alda daiteke.

6. *Fitxategi txantiloia (template, plantilla) : .etf*

ELAN proiektua behar bezala konfiguratu delarik, konfigurazioa gorde daiteke, hau da, *lerroen* eta *mota linguistiko* parametroak gordetzen dira beste proiektu baten berriz erabiltzeko. Hortarako, joan *Fitxategia* menura (Artxivo, File) --> *Eredu bezala gorde* (Guardar como plantilla, Save As Template) zapatu. Eredua gordetzeko tokia aukera ezazu eta izendatu, *.etf* fitxategia sortuko du.

ELAN fitxategi berri bat sortzean (landu behar diren fitxategiak aukeratu ondoren) txantiloia baten lerroak eta mota linguistikoak *Lerro* eta *Mota* menuetatik inporta daitezke.

7. *Mota Linguistiko*en inportazioa

Mota linguistiko bat inportatzeko *Mota* (Type, Tipo) menuan klik egin eta *Inportatu Mota* (Import Type, Importar tipo) elkarrizketa-kuadroa eskeiniko du, non dozierra bilatzeko *Arakatu* (Browse, Navegar) eremua zapatuko den landu behar den *fitxategi txantiloia (.etf)* inportatzeko; --> *Inportatu* (Importar, Import) balidatu --> *Itxi* (Cerrar, Close).

Nahiz eta ez diren ikusten, *Mota linguistiko*ak inportatuak izan dira. Egiaztatzeko aski da *Mota* (Tipo, Type) --> *Mota linguistiko* aldatu --> *Itxi*, egitea.

8. *Lerro*en inportazioa

Lerro bat inportatzeko *Lerro* (Tier, Línea) menuan klik egin eta *Inportatu Tira* (Import Tiers, Importar Tiras) elkarrizketa kuadroa eskainiko du, non dossier bilatzeko *Arakatu*

(*Browse, Navegar*) eremua zapatuko den inportatu behar den fitxategi txantiloia (.etf) inportatzeko; --> *Inportatu* (Importar, Import) balidatu --> *Itxi* (Cerrar, Close).

9. Etiketaren sortze eta kudeantzea

Mota linguistikoak eta lerroak definitu ondoren, partizio-eremua prest da etiketatze lanari ekiteko. Segmentaziorako multimedia erreproduzigailu eta nabigazio-barrez baliatzen da.

9.1. Multimedia erreproduzigailu eta nabigazio-barrak

Multimedia erreproduzigailu eta nabigazio barra hiru zatitan agertzen da: 1tik 11rako botoiek seinale osoa kontuan hartzen dute; 12tik 18rakoak, hautaturiko seinale zatia; eta azken bi kuadrotxoak seinalearen entzuteko moldeari dagozkio.



6 Irudia: Multimedia erreproduzigailu eta nabigazio barrak

- | | |
|-----------------------------------------|------------------------------------------------------|
| 1. Soinuaren hastapenera joateko | 13. Hautapena ezabatzeko |
| 2. Aitzineko kursorera joateko | 14. Kurtsoa hautapenaren esker/eskuin plazaratzeko |
| 3. Segundu bat lehenago plazatu denbora | 15. Aitzineko iruzkinera joateko(horizontalki) |
| 4. Aitzineko markora joateko | 16. Hurrengo iruzkinera joateko (horizontalki) |
| 5. Aitzineko pixelera joateko | 17. Aitzineko iruzkinera (bertikalki) |
| 6. Irakurri/pausa soinua + bideoa | 18. Hurrengo iruzkinera (bertikalki) |
| 7. Hurrengo pixelera joateko | 19. Hautu moldea (Mode selection, Modo selección) = |
| 8. Hurrengo markora joateko | laukia markatzen denean, irakurtzen du |
| 9. Segundu bat geroago plazatu denbora | 20. Errepika moldea (Mode boucle, Modo bucle) = |
| 10. Hurrengo kursorera joateko | laukia markatzen denean, lan partizioan hautatua den |
| 11. Soinuaren azkenera joateko | segmentua nahi den beste errepikatzen da. |
| 12. Hautapena irakurtzeko | |

9.2. Etiketak txertatzea

Lehenik, egiaztatu behar da etiketa sortzeko den lerroaren hautapena egina denez. Lerroaren izena gorritz azpimarratua egon behar du.

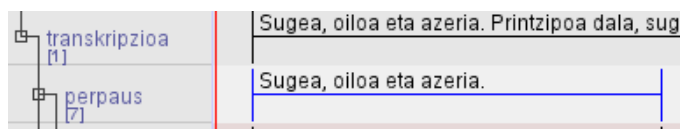
Transkribatu izan behar den eduki testualari dagokion denbora bloka hautatzen da kurtsoa gorria seinalearen gainean arrastatuz -hautatu zatia urdindua agertuko da-.

Iruzkina molde desberdinez txerta daiteke:

- *Iruzkin* (Annotation, Comentario) menutik --> *Iruzkin berria hemen* (Nouvelle annotation ici) zapatu.
- Laster menutik --> kurtsoa behar den lerroaren gainean paratuz, eskuineko botoinarekin klik egin(edo ctrl+klik mac-an) eta *Iruzkin berria hemen* klikatu.
- Alt GR + N lasterbidetik.

Idatzi nahi den iruzkina jo ondoren, *Ctrl + Enter* balidatu (*sagar + Enter* Mac-an).

Sartu (Enter) teklaz bakarrik balidatzen bada, eduki testuala ez da gordea izango eta denbora blokean lerro bat baino ez da besterik agertuko.



7 Irudia: Etiketa balidatu ondotik lortu den emaitza

Etiketa bat sortua delarik eta bere edukia edo denbora aldatu nahi bada, erraz egiten da. Bai edukia bai denboraren murriztea ala emendatzea.

10. Gordetzea

ELANen jatorrizko formatua *.eaf* da. ELANen egindako lehen gordeketa, programak sortutako fitxategiari izen bat ematea eskatuko du. Dokumentua lehenbizi gordetzeko *Fitxategia (Fichier, Archivo)* menura joan --> *Gorde honela (Enregistrer sous, Guardar como)* aukeraturaz. Adierazi behar zaio ere zein dosierretan nahi den gorde (hoberen ikus-entzunezko fitxategiak gordeta diren dosierrean).

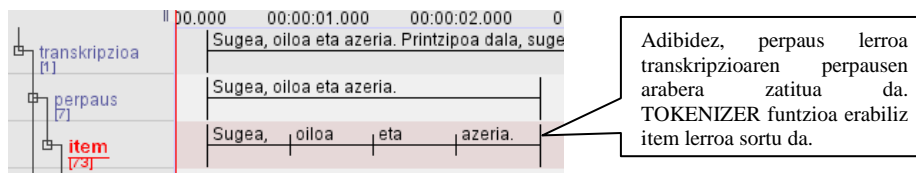
Hurrengo gordeketa, edo *Fitxategia* menutik --> *Gorde*, edo *Ctrl+S* (pc) edo *sagar+S* (mac) zapatuz egin daiteke.

Gordeketa automatikoa ere bada: *Fitxategia* menua --> *Gordeketa automatikoan (Enregistrement automatique, Copia de seguridad automática)*, zenbat minutu oro gorde behar duen adieraz daiteke: 1, 5, 10... minutu guztietan.

Gorde ondoren dokumentua hetsi dezakegu beste egun baten segitzeko. Oraingoan *.eaf* fitxategia bakarrik irekiz, ELANek automatikoki irekiko ditu dagozkion bideo eta audio fitxategiak.

11. Tokenizazioa

Funtzio honek lerro bateko unitateak bereizteko balio du. Lerroa item beste segmentutan pikatuko du. Segmentu horiek sortu behar den beste lerro baten zatiak izanen dira.



8 Irudia: Tokenizazioa

12. Inportazioa

ELANek beste programa batzuk sorturiko fitxategiak inportatzeko ahalbidea ematen du. Formatu hauek inporta daitezke:

- Shoebox
- Toolbox
- Textgrid fitxategiak, UTF-8 eta UTF-16 kodeetan
- Chat (CLAN) (.cha bakarrik): ELANek CHAT fitxategi zaharrak jasaiten ditu, baina ez XML CHAT fitxategiak
- Transcriber fitxategiak
- CSV (tabulazioz mugaturiko testu fitxategiak) (.txt edo CSV)

Halere, inportazio motaren arabera parametro desberdinak berrikusi behar dira. Emaitzak ez dira batzuetan guztiz fidelak jatorrizko fitxategiarekin. Beraz, transkripzioaren lerrokatzea ELANen pean dagoelarik berrikustea komeni da.

13. Esportazioa

Gainera ELANekin lerrokatutako transkripzioak beste formatu batzuetara esporta daitezke, beste programa batzuekin baliatzeko. Formatu hauetara esporta daiteke:

- Shoebox
- Toolbox
- CHAT
- Tiger-XML: Synpathy-ra ere esporta daiteke (enuntziatuen egitura sintaktikoen analisia eta editaia grafikoa egiteko erabiltzen den lanabesa). Tresna hau Psikolinguistikako Max Planck Institutuak garatua izan da ere.
- Lerroarte testura
- HTML
- Tabulazioz mugaturiko testura (EXCEL)
- Testu tradizionalaren transkripziara
- Hitz zerrenda
- SMIL fitxategia: Synchronized Multimedia Integration Language (SMIL) ikus-estzunezko aurkezpen interaktiboak egiteko
- Quicktime testura
- TextGrid Praat

The screenshot displays the ELAN interface with three audio channels at the top. Below the waveforms is a spectrogram. At the bottom, a TextGrid table is shown with the following content:

Sugea, oiloa eta azeria. Printzipioa dala, sugeak azeria jaten du, azeriak oiloa eta oiloak sugea. Eta egin ditugu zir												
Sugea, oiloa eta azeria.				Printzipio dala, sugeak azeria jaten du, azeriak oiloa eta oiloak sugea.								
Sugea,	oiloa	eta	azeria.	Printzi	dala,	sugeak	azeria	jaten	du,	azeriak	oiloa	eta
				PRED	V	SUBJ	OBJ	V	AUX	SUBJ	OBJ	COP
La serpiente, la gallina y el zorro. Siendo el principio, la serpiente se come al zorro, el zorro se come a la gallina												

7 280000

9 Irudia: ELAN-eko lerroakTextGrid-era esportaturik

- Azpigitulu testua

- ELAN leihotik egin irudia.

Ondorioa

Ikus-entzunezko baliabideetatik egiten diren iruzkinen hizkuntz-corpusei dimentsio berri bat ekartzen diete, multimodalitatearena.

Ez da gehiago ari solaskide baten edo gehiagoren hizketa transkribatzen eta etiketatzen bakarrik. Ikusgarri geratzen den informazioak ere interesa du: gorputz, esku edo begiraden posizio eta mugimendua, hiztunen arteko harremana, egoera eta eszenen deskripzioa.

Posibilitate hauek uzten dute hizkera bere dimentsio guztietan kodetzen, jestuak barne, zeinu hizkuntzen eta jestu ko-berbalen ikerketei bidea zabalduz.

Bibliografia

- Braffort, A., Boutora, L., Sérasset, G. (Eds.). (2012) Segmentatio du geste, Actes de la conférence conjointe JEP-TALN-RECITAL 2012. Atelier DEGELS 2012: Défi GEste Langue des Signes. <http://aclweb.org/anthology-new/W/W12/W12-12.pdf>
- Birgit, H. 2012. Manual ELAN - Linguistic Annotator version 4.4.0 <http://tla.mpi.nl/tools/tla-tools/elan/>
- Cosnier, J., Vaysse, J., (1997). Sémiotique des gestes communicatifs. In Nouveaux actes sémiotiques, 52, 7-28.
- Goodwin, C. 1979, « The interactive construction of a sentence in natural conversation ». In Everyday Language. Studies in Ethnomethodology, G. Psathas ed. 97-121. New York. Irvington Publishers.
- 1981. Conversational Organization. Interaction Between Speakers and Hearers. New York. Academic Press.
- & Goodwin, M.H. 1986. « Gesture and Coparticipation in the Activity of Searching for Word ». Semiotica. 621-2 : 51-75.
- 2000. « Action and embodiment within situated human interaction ». Journal of Pragmatics. 32 : 1489-1522.
- Mondada, L. 2003. Observer les activités de la classe dans leur diversité: choix méthodologiques et enjeux théoriques. In Actes du Colloque Ensenyar llengües en la diversitat i per a la diversitat (Barcelona, 30.5-1.6.2002) (pp. 49-70). Barcelona: ICE Universitat de Barcelona.
- 2004. “Temporalité, séquentialité et multimodalité au fondement de l’organisation de l’interaction. Le pointage comme pratique de prise du tour”. Cahiers de Linguistique Française. 26 : 169-192.
- 2005a. L’analyse de corpus en linguistique interactionnelle : de l’étude de cas singuliers à l’étude de collections. In A. Condamine (Ed.), Sémantique et corpus (pp. 76-108). Paris: Hermès.
- 2005b. La constitution de l’origo déictique comme travail interactionnel des participants. Intellectica, 2/3(41-42), 75-100.
- 2006a. Video recording as the reflexive preservation-configuration of phenomenal features for analysis. In H. Knoblauch, J. Raab, H.-G. Soeffner & B. Schnettler (Eds.), Video Analysis. Bern: Lang.

- 2006b. “La pertinenza del dettaglio, registrazione e trascrizione di dati video per la linguistica interazionale”. In Y. Bürki and E. de Stefani eds. *Transcrivere la lingua. Dalla filologia all’analisi conversazionale*. 313-44. Bern. Lang
 - 2007. “Multimodal resources for turn-taking. Pointing and the emergence of possible next speakers”. *Discourse Studies*. 9-2 : 195-226.
 - 2008a. “La transcription dans la perspective de la linguistique interactionnelle” et “Documenter l’articulation des ressources multimodales dans le temps : la transcription d’enregistrements vidéos d’interactions”. In M. Bilger (Ed.), *Données orales, les enjeux de la transcription* (pp. 78-109 et 127-155). Perpignan: Presses Universitaires de Perpignan.
 - 2008b. L’analyse de collections de phénomènes multimodaux en linguistique interactionnelle. *Cahiers de Praxématique*, (50), 23-70.
 - 2009. Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics*, 41, 1977-1997.
 - 2011. Constitution et exploitation de corpus vidéo en linguistique interactionnelle : rendre disponibles les détails multimodaux pertinents pour l’activité située. Actes du VIIe Colloque Jeunes Chercheurs Praxiling.
- Zouinar, M., Relieu, M., Salembier, P. (2004). *Observation et capture de données sur l’interaction multimodale en mobilité*. *Mobilité & Ubiquité’04*, Nice: ACM.