

O corpus ParlaMint de discursos parlamentarios

Un recurso lingüístico
para a investigación en humanidades

Adina Ioana Vladu

Instituto da Lingua Galega | Proxecto Nós

2 de maio de 2024 | Xeira CLARIAH-GAL

ParlaMint



O proxecto ParlaMint



O PROXECTO ParlaMint

Proxecto clave de CLARIN:

- financiado por CLARIN ERIC entre 2020–2023
- integra membros de 20+ nodos CLARIN

O *dataset*:

- transcricións de debates parlamentarios de 29 parlamentos nacionais e rexionais, con codificación homoxénea
- cobertura entre 2015–2022 (mínimo)
- enriquecido con metadatos, anotación lingüística e recoñecemento de entidades nomeadas
- con tradución automática ao inglés

Principais logros:

- resolve o problema da falta de homoxeneidade no formato dos datos parlamentarios europeos
- Ofrece recursos:
 - visibles
 - interoperables
 - comparables
 - accesibles

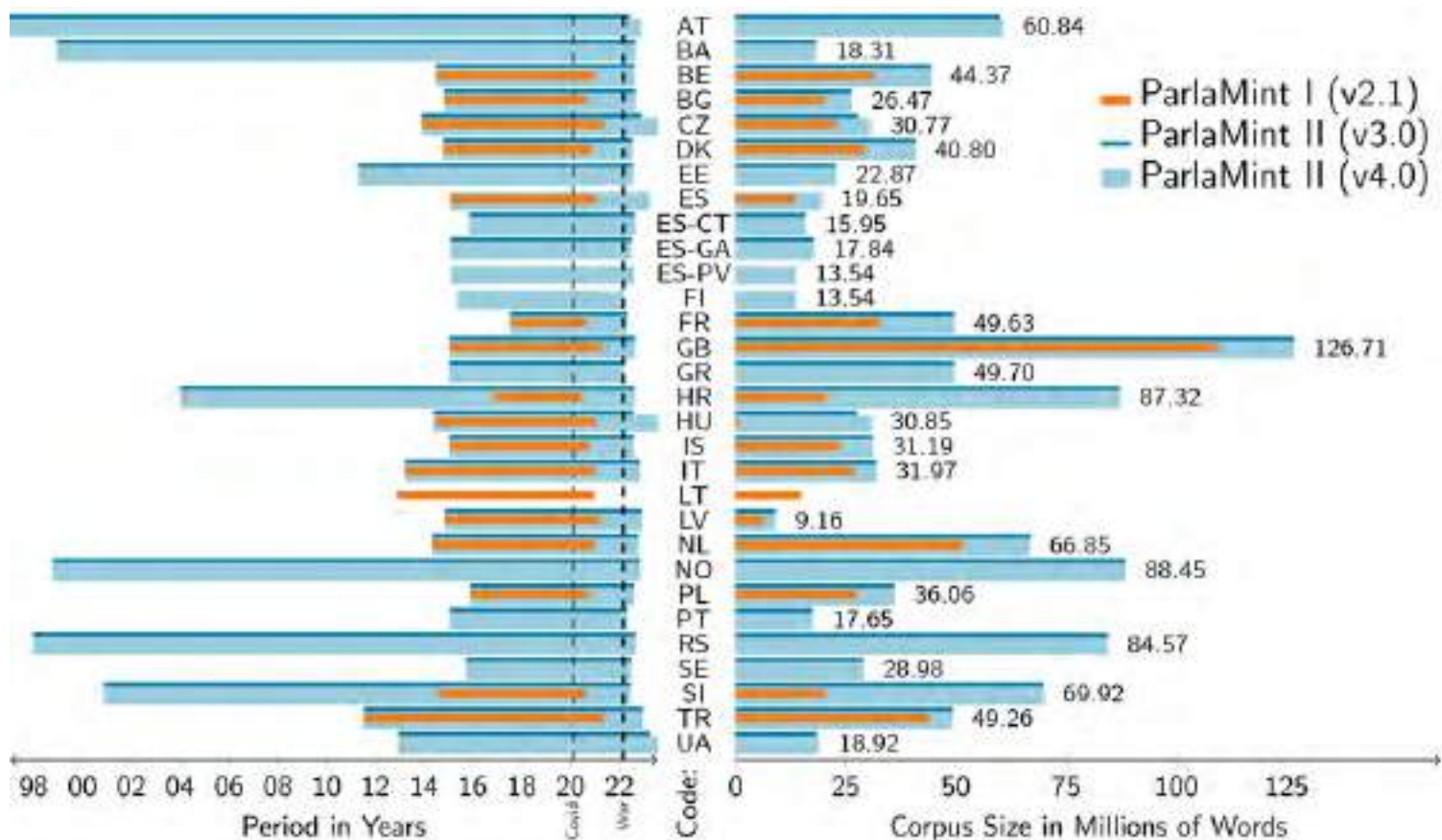
MEMBROS: países e rexións

Austria
País Vasco
Bosnia e Herzegovina
Bélxica
Bulgaria
Cataluña
Croacia
República Checa
Dinamarca
Estonia
Finlandia
Francia
Galicia
Grecia
Hungría

Islandia
Italia
Letonia
Países Baixos
Noruega
Polonia
Portugal
Serbia
Eslovenia
España
Suecia
Turquía
Ucraina



COBERTURA: duración e volume dos datos



Fonte dos datos e división en subcorpus

Debates parlamentarios:

- transcricións manuais ou semiautomáticas revisadas por persoal dos respectivos Parlamentos
- libremente accesibles nas páxinas web dos Parlamentos nacionais e rexionais

Subcorpus:

- **REFERENCIA:** ata o 30.01.2020
- **COVID:** desde o 31-01-2021, cando a OMS fixo a declaración oficial da Emerxencia de saúde pública de importancia internacional (PHEIC, polas súas siglas en inglés)
- **GUERRA:** desde o 24-02-2022 (a invasión de Ucraína por parte de Rusia)

METADATOS

- **Metadatos:** información extraída de fontes parlamentarias e de fontes externas (manual ou semiautomáticamente)
 - **Parlamento e sesións:** tipo de parlamento (uni/bicameral); número e datas das lexislaturas e sesións;
 - **Falantes:** nome, xénero, rol (MP/MG), afiliación política;
 - para algúns corpus: datos biográficos; ligazóns á Wikipedia; pertenza a diferentes comisións etc.;
 - **Partidos:** posición no goberno (coalición/oposición), orientación política (D-E), [variables CHES](#) (postura ideolóxica);
 - **Intervencións:**
 - nome do/a falante e o seu papel no debate (“chair”, “regular speaker”, “guest”);
 - márcanse e clasifícanse as notas de transcripción (p.ex., interrupcións, aplausos, murmurios etc.).

ANNOTACIÓN

Anotación lingüística automática:

- tokenización, lematización, segmentación de frases, anotación morfosintáctica;
- formato Universal Dependencies

é que eu non **gasto** os meus tempos de intervención

gasto	
UD POS	VERB
Tag	
Features	Mood=Ind, Number=Sing, Person=1, Tense=Pres
Lemma	gastar

Recoñecemento de entidades nomeadas:

- categorías: PER, LOC, ORG e MISC

Anotación semántica:

- sobre a versión con tradución automática ao inglés (ParlaMint-XX-en) (formato USAS)

a **Cámara do Parlamento**, deamos resposta ás preocupacións

Name	Cámara do Parlamento
type	Organization

ParlaMint^{MT}

Tradución automática ao inglés:

- modelos preadestrados OPUS-MT
 - baseados en arquitecturas Transformer e MarianNMT;
 - adestrados con corpus paralelos do repositorio OPUS;
 - dispoñibles para todas as linguas de ParlaMint;
- posprocesamento para corrixir a tradución dos nomes propios.

The screenshot displays the CONCORDANCE interface for ParlaMint-XX-en 4.0. At the top, it shows search statistics: 'simple: vaccines • 71,300' and 'shuffle: 71,300' with a p-value of 0.0042%. Below this, there are navigation icons and a list of search results. Each result entry includes a checkbox, a language code (e.g., GB-en, CZ-en, BE-en), a date, and a snippet of text with the word 'vaccines' highlighted in red. The interface also features a sidebar with various tool icons and a search bar at the top right.

INTEROPERABILIDADE E COMPARABILIDADE: Logros

Os corpus ParlaMint son *interoperables*:

- empregan un esquema común de anotación baseado nas guías TEI.

Os corpus ParlaMint son *comparables*:

- abarcan o mesmo período temporal: 2015–2022;
- empregan os mesmos metadatos;
- fan uso do mesmo tipo de anotación lingüística automática.

A tradución automática ao inglés aumenta a súa interoperabilidade e comparabilidade.

INTEROPERABILIDADE E COMPARABILIDADE: Retos

- Os corpus parlamentarios reflicten as especificidades de cada parlamento, o que pode significar diferenzas nos datos.
- Ás veces os fenómenos similares poden tratarse de formas diferentes.
- Algúns metadatos poden ser máis ricos/pobres, ou incluso faltar, dependendo do difícil que sexa obtelos.
- Pode haber diferenzas na precisión e os esquemas da anotación lingüística automática.

MAPA do proxecto ParlaMint

- **Guías Parla-CLARIN:**
<https://clarin-eric.github.io/parla-clarin/>
- **Mostras dos corpus incluídos en ParlaMint:**
<https://github.com/clarin-eric/ParlaMint>
- **GitHub** como plataforma colaborativa:
<https://github.com/clarin-eric/ParlaMint>
- **Repositorios para descarga dos corpus finais:**
<http://hdl.handle.net/11356/1860> (corpus anotados)
<http://hdl.handle.net/11356/1859> (corpus non anotados)
- **Ferramentas de consulta en liña:**
[noSketchEngine](#); [KonText](#); [TEITOK](#)

ParlaMint-ES-GA



ParlaMint



O corpus ParlaMint galego

(ParlaMint-ES-GA)

- **Participantes: ILG, Proxecto Nós (USC)**
 - Adina Vladu
 - Carmen Magariños
 - Daniel Bardanca
 - Elisa Fernández Rei
- **Datos:**
 - **Datos fonte:** 302 sesións plenarias (2015-2022)
 - **Metadatos:** 227 interventores, partidos políticos, GPs, legislaturas e gobernos.
- **Procesamento automático de datos:**
 - **Metadatos** -> documento XML-TEI
 - **Diarios de sesións** -> conversión TXT-TEI:
 - Identificación de interventores e roles
 - Etiquetado de intervencións (inicio e fin)
 - Segmentación por parágrafos
 - Etiquetado de notas de transcripción
- **Anotación lingüística automática**
 - Mario Barcala (NLPGO)
 - Modelo NER: Marcos Garcia (USC)



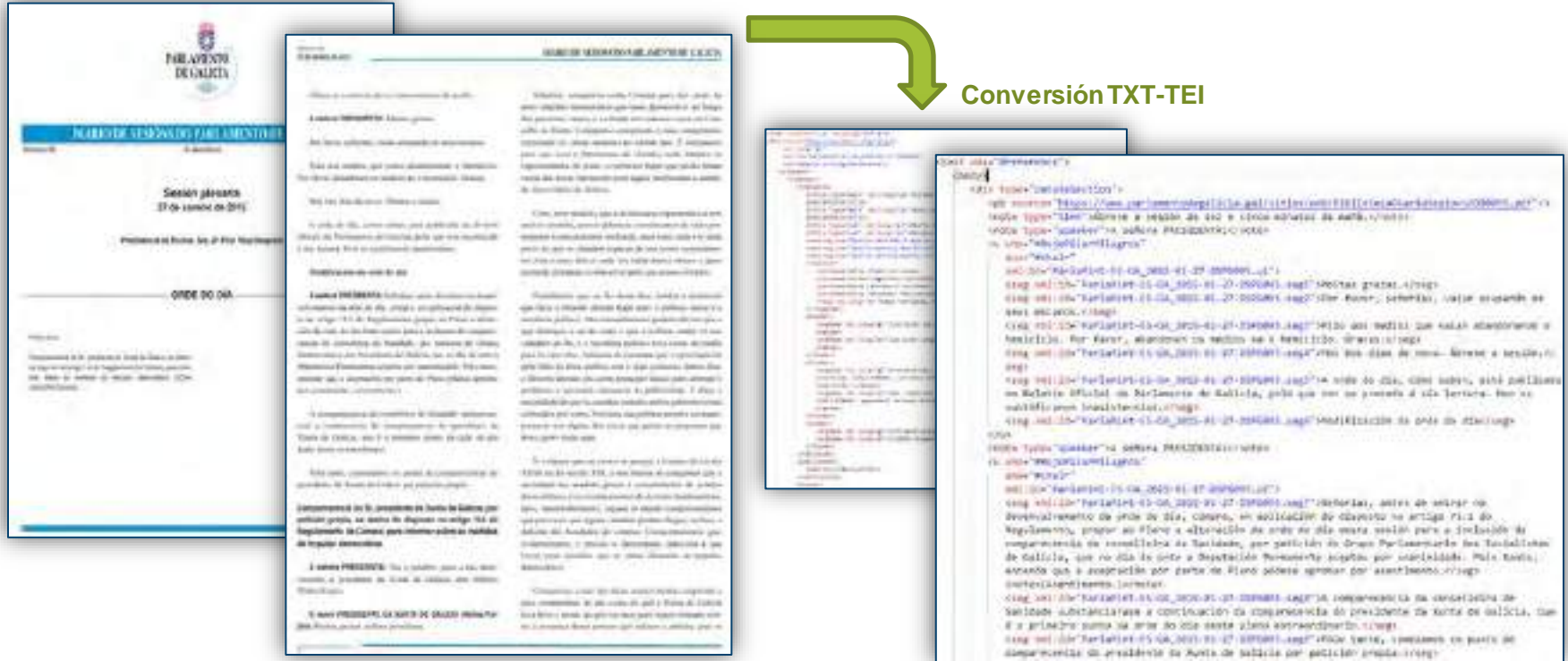
Falantes

```
<?xml version="1.0" encoding="UTF-8"?>
<listPerson xmlns="http://www.tei-c.org/ns/1.0" xml:id="ParlaMint-ES-GA-listPerson" xml:lang="gl">
  <person xml:id="LorenzanaMaríaJesús">
    <persName>
      <surname>Lorenzana</surname>
      <surname>Somoza</surname>
      <forename>María Jesús</forename>
    </persName>
    <sex value="F"/>
    <birth when="1981">
      <placeName>A Coruña</placeName>
    </birth>
    <affiliation role="member" ref="#PG" from="2020-09-01" ana="#PG.11"/>
    <affiliation role="minister" ref="#GOV" from="2020-09-07" ana="#GOV.11">
      <roleName>Conselleira de Emprego e Igualdade</roleName>
    </affiliation>
    <affiliation role="member" ref="#GOV" from="2020-09-07" ana="#GOV.11"/>
    <idno type="URI" subtype="wikimedia">https://gl.wikipedia.org/wiki/Mar%C3%ADa\_Xes%C3%BAs\_Lorenzana</idno>
  </person>
```

Partidos políticos

```
<org xml:id="party.BNG" role="politicalParty">
  <orgName full="yes" xml:lang="gl">Bloque Nacionalista Galego</orgName>
  <orgName full="yes" xml:lang="en">Galician Nationalist Bloc</orgName>
  <orgName full="abb">BNG</orgName>
  <event from="1982-09-15">
    <label xml:lang="en">existence</label>
  </event>
  <idno type="URI" xml:lang="gl" subtype="wikimedia">https://gl.wikipedia.org/wiki/Bloque\_Nacionalista\_Galego</idno>
  <idno type="URI" xml:lang="en" subtype="wikimedia">https://en.wikipedia.org/wiki/Galician\_Nationalist\_Bloc</idno>
</org>
<org xml:id="party.PPdeG" role="politicalParty">
  <orgName full="yes" xml:lang="gl">Partido Popular de Galicia</orgName>
  <orgName full="yes" xml:lang="en">People's Party of Galicia</orgName>
  <orgName full="abb">PPdeG</orgName>
  <event from="1989-01-12">
    <label xml:lang="en">existence</label>
  </event>
  <idno type="URI" xml:lang="gl" subtype="wikimedia">https://gl.wikipedia.org/wiki/Partido\_Popular\_de\_Galicia</idno>
  <idno type="URI" xml:lang="en" subtype="wikimedia">https://en.wikipedia.org/wiki/People%27s\_Party\_of\_Galicia</idno>
</org>
<org xml:id="party.PSdeG-PSOE" role="politicalParty">
  <orgName full="yes" xml:lang="gl">Partido de los Socialistas de Galicia</orgName>
  <orgName full="yes" xml:lang="en">Socialists' Party of Galicia</orgName>
  <orgName full="abb">PSdeG-PSOE</orgName>
  <event from="1979-05-02">
    <label xml:lang="en">existence</label>
  </event>
  <idno type="URI" xml:lang="gl" subtype="wikimedia">https://gl.wikipedia.org/wiki/Partido\_dos\_Socialistas\_de\_Galicia</idno>
  <idno type="URI" xml:lang="en" subtype="wikimedia">https://en.wikipedia.org/wiki/Socialists%27\_Party\_of\_Galicia</idno>
</org>
```


Dos diarios de sesión ao formato TEI



Diarios de sesión en formato TEI (exemplo)

A señora PRESIDENTA: Grupo Mixto, señora Iglesias e señora Martínez.

A señora IGLESIAS SUEIRO: Bos días, señora conselleira; bos días a todas. Grazas, presidenta.

De todas é coñecido o caso rampante nas Urxencias nos últimos días, que coincide co andado da gripe: espera de ata sete horas, profesionais desbordadas, escaseza de camas e de material...

E pese á petición de perdón por parte do presidente, como se dun pecado se tratare e non dacha responsabilidade política, e dos argumentos desta conselleira, que reduce a situación caótica a un problema de confort, señora conselleira, -señora conselleira, ¡que falta de respecto!—

A señora PRESIDENTA: Señora Iglesias.

A señora IGLESIAS SUEIRO: ...e das poucas disminaladas intencións de culpabilizar as pacientes (situarlas), pois parece que enfermamos por riba das nosas posibilidades...



Conversión TXT-TEI

```
1048 <note type="speaker">A señora PRESIDENTA:</note>
1049 <cu who="MdoPilarIglesias" ana="schain" xml:base="ParlaMint-ES-GA_2015-01-27-DSP000.v64">
1050 <seg xml:base="ParlaMint-ES-GA_2015-01-27-DSP000.seg417">Grupo Mixto, señora Iglesias e señora Martínez.</seg>
1051 </cu>
1052 <note type="speaker">A señora IGLESIAS SUEIRO:</note>
1053 <cu who="MdoPilarIglesias" ana="wregular" xml:base="ParlaMint-ES-GA_2015-01-27-DSP000.v64">
1054 <seg xml:base="ParlaMint-ES-GA_2015-01-27-DSP000.seg418">Bos días, señora conselleira; bos días a todas. Grazas,
1055 presidenta.</seg>
1056 <seg xml:base="ParlaMint-ES-GA_2015-01-27-DSP000.seg419">De todas é coñecido o caso rampante nas Urxencias nos
1057 últimos días, que coincide co andado da gripe: espera de ata sete horas, profesionais desbordadas, escaseza de camas e de
1058 material...</seg>
1059 <seg xml:base="ParlaMint-ES-GA_2015-01-27-DSP000.seg420">E pese á petición de perdón por parte do presidente,
1060 como se dun pecado se tratare e non dacha responsabilidade política, e dos argumentos desta conselleira, que reduce a
1061 situación caótica a un problema de confort, señora conselleira, -señora conselleira, ¡que falta de respecto!—</seg>
1062 </cu>
1063 <note type="speaker">A señora PRESIDENTA:</note>
1064 <cu who="MdoPilarIglesias" ana="schain" xml:base="ParlaMint-ES-GA_2015-01-27-DSP000.v64">
1065 <seg xml:base="ParlaMint-ES-GA_2015-01-27-DSP000.seg421">Señora Iglesias.</seg>
1066 </cu>
1067 <note type="speaker">A señora IGLESIAS SUEIRO:</note>
1068 <cu who="MdoPilarIglesias" ana="wregular" xml:base="ParlaMint-ES-GA_2015-01-27-DSP000.v64">
1069 <seg xml:base="ParlaMint-ES-GA_2015-01-27-DSP000.seg422">...e das poucas disminaladas intencións de culpabilizar as
1070 pacientes (situarlas), pois parece que enfermamos por riba das nosas posibilidades...
1071 <sup type="murmuring">
1072 <del>{Murmuring}</del>
1073 </sup>
1074 </cu>
1075 <seg>
1076 , pois parece que enfermamos por riba das nosas posibilidades...
1077 </seg>
1078 </cu>
```

Identificación de falantes

```
1854 <note type="speaker">A señora PRESIDENTA:</note>
1855 <u who="Rojo Pilar Iglesias" ana="chair" xml:id="ParlaMint-ES-GA_2015-01-27-DSP0095.u04">
1856 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP0095.seg457">Grupo Mixto, señora Iglesias e señora Martínez.</seg>
1857 </u>
1858 <note type="speaker">A señora IGLESIAS SUEIRO:</note>
1859 <u who="Iglesias Carmen" ana="regular" xml:id="ParlaMint-ES-GA_2015-01-27-DSP0095.u05">
1860 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP0095.seg458">Bos días, señora conselleira; bos días a todas, óras,
1861 presidenta.</seg>
1862 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP0095.seg459">De todas é coñecido o caos rampante nas urxencias nas
1863 últimas datas, que coinciden co andazo da gripe: esperas de ata sete horas, profesionais desbordadas, escaseza de camas e de
1864 material...</seg>
1865 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP0095.seg460">E pese á petición de perdón por parte do presidente,
1866 coma se dun pecado se tratara e non durma responsabilidade política, e dos argumentos desta conselleira, que reduce a
1867 situación cábica a un problema de confort, señora conselleira, -señora conselleira, (que falta de respecto!...</seg>
1868 </u>
1869 <note type="speaker">A señora PRESIDENTA:</note>
1870 <u who="Rojo Pilar Iglesias" ana="chair" xml:id="ParlaMint-ES-GA_2015-01-27-DSP0095.u06">
1871 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP0095.seg461">Señora Iglesias.</seg>
1872 </u>
1873 <note type="speaker">A señora IGLESIAS SUEIRO:</note>
1874 <u who="Iglesias Carmen" ana="regular" xml:id="ParlaMint-ES-GA_2015-01-27-DSP0095.u07">
1875 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP0095.seg462">
1876 ...e das poucas disimuladas intencións de culpabilizar as pacientes
1877 <vocal type="murmuring">
1878 <desc>(Murmurios.)</desc>
1879 </vocal>
1880 , pois parece que enfermamos por riba das nosas posibilidades...
1881 </seg>
1882 </u>
```

Identificación de rol

```
1054 <note type="speaker">A señora PRESIDENTA:</note>
1055 <u who="#RojoPilarMilagro" ana="#chair" xml:id="ParlaMint-ES-GA_2015-01-27-DSP005.u04">
1056 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP005.seg457">Grupo Mixto, señora Iglesias e señora Martínez.</seg>
1057 </u>
1058 <note type="speaker">A señora IGLESIAS SUEIRO:</note>
1059 <u who="#IglesiasCarmen" ana="#regular" xml:id="ParlaMint-ES-GA_2015-01-27-DSP005.u05">
1060 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP005.seg458">Bos días, señora conselleira; bos días a todas. Grazas,
presidenta.</seg>
1061 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP005.seg459">De todas é coñecido o caos rampante nas Urxencias nas
últimas datas, que coinciden co andazo da gripe: esperas de ata sete horas, profesionais desbordadas, escaseza de camas e de
material...</seg>
1062 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP005.seg460">É pese á petición de perdón por parte do presidente,
cosa se dun pecado se tratara e non dunha responsabilidade política, e dos argumentos desta conselleira, que reduce a
situación caótica a un problema de confort, señora conselleira, -señora conselleira, ¡que falta de respecto!...</seg>
1063 </u>
1064 <note type="speaker">A señora PRESIDENTA:</note>
1065 <u who="#RojoPilarMilagro" ana="#chair" xml:id="ParlaMint-ES-GA_2015-01-27-DSP005.u06">
1066 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP005.seg461">Señora Iglesias.</seg>
1067 </u>
1068 <note type="speaker">A señora IGLESIAS SUEIRO:</note>
1069 <u who="#IglesiasCarmen" ana="#regular" xml:id="ParlaMint-ES-GA_2015-01-27-DSP005.u07">
1070 <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSP005.seg462">
1071 <u>
1072 <vocal type="murmuring">
1073 <desc>(Murmurios.)</desc>
1074 </vocal>
1075 <u> pois parece que enfermamos por riba das nosas posibilidades...
1076 </seg>
1077 </u>
```

Segmentación de intervencions

1054	<note type="speaker">A señora PRESIDENTIA:</note>
1055	<u who="#RojoPilarMilagros" ana="#chair" xml:Id="ParlaMint-ES-GA_2015-01-27-DSPG095.u64">
1056	<seg xml:Id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457">Grupo Mixto, señora Iglesias e señora Martínez.</seg>
1057	</u>
1058	<note type="speaker">A señora IGLESIAS SUEIRO:</note>
1059	<u who="#IglesiasCarmen" ana="#regular" xml:Id="ParlaMint-ES-GA_2015-01-27-DSPG095.u65">
1060	<seg xml:Id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg458">Bos días, señora conselleira; bos días a todas. Grazas, presidenta:</seg>
1061	<seg xml:Id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg459">De todas é coñecido o caos rampante nas Urxencias nas últimas dotas, que coinciden co andar da gripe: esperas de ata sete horas, profesionais desbordadas, escaseza de camas e de material...</seg>
1062	<seg xml:Id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg460">E pasa á petición de perdón por parte do presidente, coma se dun pecado se tratare e non dunha responsabilidade política, e dos argumentos desta conselleira, que reduce a situación caótica a un problema de confort, señora conselleira, -señora conselleira, ¡que falta de respecto!...</seg>
1063	</u>
1064	<note type="speaker">A señora PRESIDENTIA:</note>
1065	<u who="#RojoPilarMilagros" ana="#chair" xml:Id="ParlaMint-ES-GA_2015-01-27-DSPG095.u66">
1066	<seg xml:Id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg461">Señora Iglesias.</seg>
1067	</u>
1068	<note type="speaker">A señora IGLESIAS SUEIRO:</note>
1069	<u who="#IglesiasCarmen" ana="#regular" xml:Id="ParlaMint-ES-GA_2015-01-27-DSPG095.u67">
1070	<seg xml:Id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg462">
1071	...e das poucas disimuladas intencións de culpabilizar as pacientes
1072	<vocal type="murmuring">
1073	<desc>(Murmurios.)</desc>
1074	</vocal>
1075	, pois parece que enfermamos por riba das nosas posibilidades...
1076	</seg>
1077	</u>

Notas de transcripción

```
1054 <note type="speaker">A señora PRESIDENTA:</note>
1055 <u who="MrojoPilarMilagros" ana="Mchair" xml:id="ParlaMint-ES-GA_2015-01-27-0SPG095.u64">
1056 <seg xml:id="ParlaMint-ES-GA_2015-01-27-0SPG095.seg457">Grupo Mixto, señora Iglesias e señora Martínez.</seg>
1057 </u>
1058 <note type="speaker">A señora IGLESIAS SUEIRO:</note>
1059 <u who="IglesiasCarmen" ana="Regular" xml:id="ParlaMint-ES-GA_2015-01-27-0SPG095.u65">
1060 <seg xml:id="ParlaMint-ES-GA_2015-01-27-0SPG095.seg458">Bos días, señora conselleira; bos días a todas. Grazas,
presidenta.</seg>
1061 <seg xml:id="ParlaMint-ES-GA_2015-01-27-0SPG095.seg459">De todas á coñecido o caos rampante nas Urxencias nas
últimas datas, que coñecen co andazo da gripe: espesas de ata sete horas, profesionais desbordadas, escaseza de camas e de
material...</seg>
1062 <seg xml:id="ParlaMint-ES-GA_2015-01-27-0SPG095.seg460">E pese á petición de perdón por parte do presidente,
coma se dun pecado se tratare e non dunha responsabilidade política, e dos argumentos desta conselleira, que reduce a
situación caótica a un problema de confort, señora conselleira, -señora conselleira, ¡que falta de respecto!...</seg>
1063 </u>
1064 <note type="speaker">A señora PRESIDENTA:</note>
1065 <u who="MrojoPilarMilagros" ana="Mchair" xml:id="ParlaMint-ES-GA_2015-01-27-0SPG095.u66">
1066 <seg xml:id="ParlaMint-ES-GA_2015-01-27-0SPG095.seg461">Señora Iglesias.</seg>
1067 </u>
1068 <note type="speaker">A señora IGLESIAS SUEIRO:</note>
1069 <u who="IglesiasCarmen" ana="Regular" xml:id="ParlaMint-ES-GA_2015-01-27-0SPG095.u67">
1070 <seg xml:id="ParlaMint-ES-GA_2015-01-27-0SPG095.seg462">
...e das poucas disimuladas intencións de culpabilizar as pacientes
1071 <vocal type="murmuring">
1072 <desc>(Murmurios.)</desc>
1073 </vocal>
1074 , pois parece que enfermamos por riba das nosas posibilidades...
1075 </seg>
1076 </u>
1077 </u>
```

Anotación lingüística

- Lematización e categorías gramaticais

```
cu who="#BojopilarMilagros" ana="#chair" xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.u1">
  cseg xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg1">
    <u xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg1.s1">
      <w xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg1.s1.w1" lemma="noito" msd="UPosTag=DET|Definite=Ind|Gender=Fem|Number=Plur|PronType=Art">noitas</w>
      <w xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg1.s1.w2" lemma="graza" msd="UPosTag=ADJ|Gender=Fem|Number=Plur" join="right">grazasc</w>
      <pc xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg1.s1.u1" msd="UPosTag=PUNCT|PunctType=Peri">.</pc>
      <linktrp targFunc="head argument" type="UD-SYN">
        <link ana="ud-syn:det" target="#ParlaMint-ES-GA_2015-01-27-DSPG095.seg1.s1.w2 #ParlaMint-ES-GA_2015-01-27-DSPG095.seg1.s1.w1"/>
        <link ana="ud-syn:root" target="#ParlaMint-ES-GA_2015-01-27-DSPG095.seg1.s1 #ParlaMint-ES-GA_2015-01-27-DSPG095.seg1.s1.w1"/>
        <link ana="ud-syn:punct" target="#ParlaMint-ES-GA_2015-01-27-DSPG095.seg1.s1.u1 #ParlaMint-ES-GA_2015-01-27-DSPG095.seg1.s1.w1"/>
      </linktrp>
    </u>
  </cseg>
  cseg xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2">
    <u xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1">
      <w xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.w1" lemma="por" msd="UPosTag=ADP|AdpType=Prep">por</w>
      <w xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.w2" lemma="Favor" msd="UPosTag=NOUN|Gender=Masc|Number=Sing" join="right">Favor</w>
      <pc xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.u1" msd="UPosTag=PUNCT|PunctType=Comm">,</pc>
      <w xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.w4" lemma="señoría" msd="UPosTag=NOUN|Gender=Fem|Number=Plur" join="right">señorías</w>
      <pc xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.u5" msd="UPosTag=PUNCT|PunctType=Comm">,</pc>
      <w xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.w6" lemma="Ir" msd="UPosTag=VERB|Mood=Sub|Number=Plur|Person=3|Tense=Pres">va<sup>ia</sup>nc</w>
      <w xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.w7" lemma="ocupar" msd="UPosTag=VERB|VerbForm=Ger">ocupando</w>
      <w xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.w8" lemma="o" msd="UPosTag=DET|Definite=Def|Gender=Masc|Number=Plur|PronType=Art">os</w>
      <w xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.w9" lemma="seu" msd="UPosTag=DET|Definite=Def|Gender=Masc|Number=Plur|Person=3|Poss=Yes">seus</w>
      <w xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.w10" lemma="escano" msd="UPosTag=NOUN|Gender=Masc|Number=Plur" join="right">escanos</w>
      <pc xsl:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg2.s1.u11" msd="UPosTag=PUNCT|PunctType=Peri">.</pc>
    </u>
  </cseg>
</s>
```

Anotación lingüística

- Anotación gramatical (*Universal Dependencies*)

```
</seg>
<seg xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2">
  <s xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1">
    cw xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w1" lemma="por" msd="UPostag=ADP|NdType=Prep">por</w>
    cw xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w2" lemma="favor" msd="UPostag=NN|Gender=Masc|Number=Sing" join="right">favor</w>
    cpc xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w3" msd="UPostag=PUNCT|PunctType=Comm">,</pc>
    cw xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w4" lemma="señoría" msd="UPostag=NOUH|Gender=Fem|Number=Plur" join="right">señoría</w>
    cpc xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w5" msd="UPostag=PUNCT|PunctType=Comm">,</pc>
    cw xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w6" lemma="ir" msd="UPostag=VERB|Mood=Sub|Number=Plur|Person=3|Tense=Pres">vayan</w>
    cw xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w7" lemma="ocupar" msd="UPostag=VERB|VerbForm=Ger">ocupando</w>
    cw xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w8" lemma="o" msd="UPostag=DET|Definite=Def|Gender=Masc|Number=Plur|PronType=Art">os</w>
    cw xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w9" lemma="sea" msd="UPostag=DEF|Definite=Def|Gender=Masc|Number=Plur|Person=3|Poss=Yes">seas</w>
    cw xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w10" lemma="oscaro" msd="UPostag=NOUH|Gender=Masc|Number=Plur" join="right">oscaros</w>
    cpc xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w11" msd="UPostag=PUNCT|PunctType=Par1">,</pc>
    <linkerp targfunc="head argument" type="UD-SYN">
      <link ana="ud-syn:case" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w2 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w1"/>
      <link ana="ud-syn:obj" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w6 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w2"/>
      <link ana="ud-syn:punct" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w4 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w1"/>
      <link ana="ud-syn:conj" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w2 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w6"/>
      <link ana="ud-syn:punct" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w2 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w5"/>
      <link ana="ud-syn:root" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w6"/>
      <link ana="ud-syn:advcl" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w6 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w7"/>
      <link ana="ud-syn:det" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w8 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w9"/>
      <link ana="ud-syn:det" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w10 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w7"/>
      <link ana="ud-syn:obj" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w7 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w10"/>
      <link ana="ud-syn:punct" target="#ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w7 #ParlaMint-ES-GA_2015-01-27-DSPG095_seg2.s1.w11"/>
    </linkerp>
  </s>
</seg>
```


Anotación lingüística

- Reconocimiento de entidades nomeadas (NER)

```
<note type="speaker" xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.ana.note101">A señora PRESIDENTA:</note>
<u who="#RojoPilarMilagros" ana="#chair" xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.u64">
  <seg xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457">
    <s xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1">
      <name xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.ne1" type="ORG">
        <w xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.w1" lemma="grupo" msd="UPosTag=NCUN|Gender=Masc|Number=Sing">Grupo</w>
        <w xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.w2" lemma="mixto" msd="UPosTag=ADJ" join="right">Mixto</w>
      </name>
      <pc xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.w3" msd="UPosTag=PUNCT|PunctType=Comm">,</pc>
      <w xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.w4" lemma="señor" msd="UPosTag=ADJ">señora</w>
      <name xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.ne3" type="PER">
        <w xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.w5" lemma="iglesias" msd="UPosTag=PROPN">Iglesias</w>
      </name>
      <w xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.w6" lemma="e" msd="UPosTag=CCONJ">e</w>
      <w xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.w7" lemma="señor" msd="UPosTag=NCUN|Gender=Fem|Number=Sing">señora</w>
      <name xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.ne5" type="PER">
        <w xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.w8" lemma="martínez" msd="UPosTag=PROPN" join="right">Martínez</w>
      </name>
      <pc xml:id="ParlaMint-ES-GA_2015-01-27-DSPG095.seg457.s1.w9" msd="UPosTag=PUNCT|PunctType=Peri">.</pc>
```

Acceso



ParlaMint



Acceso aos corpus:

Disponibles para consulta nas seguintes ferramentas:

1. [noSketchEngine](#) sen rexistro de usuario e [noSketchEngine](#) con rexistro de usuario
2. [KonText](#)
3. [TEITOK](#)

Disponibles para descarga desde o repositorio CLARIN.SI:

1. En formato [TEI “texto plano”](#)
2. [Con anotación lingüística](#)
3. [Traducidos ao inglés](#) (*Nota: a tradución automática pode conter erros de contido típicos da tradución automática neuronal, incluso cando se alcanza unha boa calidade.)

No repositorio [GitHub de ParlaMint](#) pódense atopar mostras dos corpus, os esquemas XML e os scripts de validación automática.

ParlaMint en TEITOK



ParlaMint

Descrición dos corpus



TEITOK

Login
Available Corpora

ParlaMint 4.0

About
Subcorpora

ParlaMint-ES-GA

Transcriptions
People
Organizations
Corpus Search

Powered by CC-BY
Maarten Janssen, 2014

Subcorpora

Corpus parlamentario galego ParlaMint-ES-GA

Source(s)

Actas do Parlamento de Galicia

27.01.2015 - 25.05.2022

Location(s)

Rúa do Hórreo, 63, 15702 Santiago de Compostela, A Coruña, Galicia, España - Santiago de Compostela - Galicia 2015-01-25 - 2022-05-25

Names database

Transcriptions

Corpus search (23M tokens)

Linguistically annotated corpus of parliamentary debates ParlaMint-ES-GA

- Autonomous region: ES-GA (Galicia)
- Languages: gl (Galician)
- Version: 4.0
- Handle: <http://hdl.handle.net/11356/1860>

Documentation

Characteristics of the national parliament

The Galician Parliament is the legislative body of the autonomous community of Galicia, located in North-western Spain. It is composed of 75 members who are elected every four years through a proportional representation system. The Parliament is responsible for enacting legislation and overseeing the actions of the regional government.

Metadatos

ParlaMint 4.0

About

Subcorpora

ParlaMint-ES-GA

Transcriptions

People

Organizations

Corpus Search

Person

Amigo Díaz, María Encarnación

Sex: Feminine

Birth date: 1957

Birth place: As Nogais

Links

wikimedia https://gl.wikipedia.org/wiki/Encarna_Amigo

Affiliation

Role	From	To	Name	Type
Member	2016-10-21	2020-02-10	Parlamento de Galicia	Parliament
Member	2016-10-21	2020-02-10	Partido Popular de Galicia	Political Party
Member				
Member				

Organization

Parlamento de Galicia (Parliament of Galicia)

Links

wikimedia https://gl.wikipedia.org/wiki/Parlamento_de_Galicia

wikimedia https://en.wikipedia.org/wiki/Parliament_of_Galicia

Events

Label	From	To
Term 9	2012-11-16	2016-08-01
Term 10	2016-10-21	2020-02-10
Term 11	2020-08-07	

Buscas por texto

ParlaMint 4.0

About

Subcorpora

ParlaMint-ES-GA

Transcriptions

People

Organizations

Corpus Search

Corpus Search

Subcorpus ParlaMint-ES-GA (change)

CQL Query:

Search

[query builder](#) | [visualize](#) | [options](#)

1291 results · ipm: 55.13 · Showing 0 - 100 · next

Tags: UD POS Tag Features Lemma

context	ir poñerlle unha vacina a unha neta sahrauí ;	Rodríguez Estévez, David
context	o descubridor da vacina da rubéola non	Rodríguez-Viño Rodríguez, Marta María
context	incorporación da vacina do pneumococo e	Rodríguez-Viño Rodríguez, Marta María
context	o Goberno central inclina a vacina da varicela no	Rodríguez-Viño Rodríguez, Marta María
context	ano 2014 da vacina contra a hepatitis C. Despois	Prado Cores, María Montserrat
context	melhora da vacina da meningite B	Torredo Quintela, Julio
context	, para que lembren algunha vacina como a da	Torredo Quintela, Julio
context	a ausencia da vacina da TPA ou	Sola Fernández, Eva
context	país onde se necesita algunha vacina determinada : Polo	Sola Fernández, Eva
context	convertera nun efecto vacina aos electorados franceses	Puy Farga, Pedro
context	deputada Rodríguez-Viño : " A vacina contra a meningite B non	Villares Naveira, Luis
context	poñecemos e iso é unha vacina , que xa estamos vacunados	Fernandez Prado, Martín

Buscas por metadatos

ParlaMint 4.0

About

Subcorpora

ParlaMint-ES-GA

Transcriptions

People

Organizations

Corpus Search

Corpus Search

Subcorpus ParlaMint-ES-GA (change)

CQL Query:

Search

[query builder](#) | [visualize](#) | [options](#)

Query Builder

Text Search	Named Entity
Form: <input type="text" value="matches"/>	Type: <input type="text" value="select"/>
UD POS Tag: <input type="text" value="select"/>	
xpos: <input type="text" value="matches"/>	Speaker
Features: <input type="text" value="expand"/>	Name: <input type="text" value="Santales Viera, Miguel Ángel"/>
Lemma: <input type="text" value="matches"/>	Role: <input type="text" value="select"/>
<input type="button" value="Add token"/>	Party name: <input type="text"/>
	Party: <input type="text" value="select"/>
	Party orientation: <input type="text" value="select"/>
	Gender: <input type="text" value="select"/>
	MP?: <input type="text" value="select"/>
	Minister?: <input type="text" value="select"/>
	Birth year: <input type="text"/>

Lectura fácil con marcas de categoría

ParlaMint 4.0

About

Subcorpora

ParlaMint-ES-GA

Transcriptions

People

Organizations

Corpus Search

Document Browser

Documents > ES-GA - Galicia (Spain)

- Subcorpus
- Body
- Term
- Meeting
- Sitting
- Sitting date
- All documents

Named Entity View

Corpus parlamentario galego ParlaMint-ES-GA, 2017-02-07-DSPG015.ana

Country	Galicia
Source date	07.02.2017
Term	X Legislatura
Meeting	Número 015
Sitting	07.02.2017
Previous	2017/ParlaMint-ES-GA_2017-02-06-DSPG014.xml
Next	2017/ParlaMint-ES-GA_2017-02-08-DSPG016.xml

Page

Abrese a sesión ás dez e dous minutos da mañá.

O señor PRESIDENTE:

Sustentado Víctor Miguel Ángel

Imos dar comezo á sesión.

Bos días. Abrese a sesión.

A orde do día está publicada no Boletín Oficial do Parlamento e non procedemos á súa lectura. Non se xusta

Proposta de alteración da orde do día

A señora PONTÓN MONDELO pide a palabra.

¿Si? ¿Para que quere a palabra, señora Pontón?

A señora PONTÓN MONDELO:

PONTÓN MONDELO, Ana Están

Moi bo día, señor presidente.

Señorías, como sabe, o artigo 73.1 establece que a orde do día do pleno pode ser alterada por acordo deste presidente ou por petición de dous grupos parlamentarios ou dunha quinta parte dos membros da Cámara. Nós grupos que estamos na oposición solicitamos unha comparecencia urxente do **Goberno** da **Xunta** en relación coas medidas que se van adoptar ante esta cuestión. Era unha comparecencia urxente na que cremos que era in

ParlaMint en noSketch Engine



ParlaMint



USO DO CORPUS EN NOSKETCH ENGINE



- <https://www.clarin.si/skelog/>
- Premer en “*Sign up now*” e crear un usuario.

EXPLORAR O CONTIDO – estrutura do corpus

The image shows a dashboard for the ParlaMint-ES-GA 4.0 (Galician parliament) corpus. At the top left, the word "DASHBOARD" is displayed. To its right is a search bar containing the text "ParlaMint-ES-GA 4.0 (Galician parliament)". A yellow box highlights a circular icon with a plus sign next to the search bar. Below the search bar, the text "PARLAMINT-ES-GA 4.0 (GALICIAN PARLIAMENT)" is shown. To the right of this text are two buttons: "CORPUS INFO" and "MANAGE CORPUS". The main area of the dashboard is divided into several tool cards:

- Concordance**: Examples of use in context
- Parallel Concordance**: Translation search
- Wordlist**: Frequency list
- Keywords**: Terminology extraction
- Trends**: Diachronic analysis, neologisms
- Text type analysis**: Statistics of the whole corpus

At the bottom of the dashboard, there is a row of icons: a target icon, a refresh icon, a list icon, the letters "NE", a letter "A", and a logo with "AR". Below these icons, a message reads: "You are using NoSketch Engine. These tools are only available in Sketch Engine".

EXPLORAR O CONTIDO – capas de anotación

ParlaMint-ES-GA 4.0 (Galician parliament) parlamint40_es_ga

Galician parliamentary corpus ParlaMint-ES-GA, 2015-2022 v4.0 [see more](#)

MANAGE CORPUS MANAGE SUBCORPORA TEXT TYPE ANALYSIS

GENERAL INFO

Language: Galician

CORPUS DESCRIPTION &

TAGSET

PROMPT WORD CHECKBOX

WORDLIST TABS

Lista das etiquetas lingüísticas

COUNTS

Tokens	20,333,757
Words	17,511,047
Sentences	780,740
Paragraphs	115,000
Documents	83,078

LEXICON SIZES

word ^f	148,126
lc	110
norm	150,492
lemma	69,253
lemma_lc	69,253
pos	81
feats	1,056
id	588
dep	877
dep_head_lemma	123,330
dep_head_pos	332
dep_head_feats	5,370
dep_head_id	5,176

TEXT TYPES

TEXT TYPE ANALYSIS

<speech> (22)	83,078
agenda . speech:agenda	1
date . speech:date	288
meeting . speech:meeting	162
Parliamentary body . speech:body	1
party_orientation . speech:party_orientation	4
party_status . speech:party_status	3
session . speech:session	1
sitting . speech:sitting	269
speaker_birth . speech:speaker_birth	47
speaker_gender . speech:speaker_gender	2
speaker_id . speech:speaker_id	216
speaker_minimal . speech:speaker_minimal	2
speaker_np . speech:speaker_np	2
speaker_name . speech:speaker_name	216
speaker_party . speech:speaker_party	0
speaker_party_name . speech:speaker_party_name	0

EXPLORAR O CONTIDO- información sobre os metadatos

DASHBOARD ParlaMint-XX 4.0 (European parliaments)

PARLAMENTOS-GA 4.0 (GALICIAN PARLIAMENT)

- Concordance
- Wordlist
- Trends
- Parallel Concordance
- Keywords
- Text type analysis**

TEXT TYPE ANALYSIS ParlaMint-XX 4.0 (European parliaments)

Structure and text types

- speech - session
- speech - sitting
- speech - speaker_left
- speech - speaker_gender
- speech - speaker_reorder
- speech - speaker_rtp
- speech - speaker_name
- speech - speaker_party**

View: Structure frequency

Scope: None (the whole corpus)

Interacts: [] [X]

analyze multiple text types

speech - speaker_party

Party	Percentage
PPSD	60.0%
PSD-S-PRICE	12%
DNG	4.5%
CM	3.9%
AGE	3.0%
Other	1.7%

ParlaMint-XX 4.0 (European parliaments)

Items: 41, Total frequency: 7,051,128

speech - party_orientation

Orientation	Percentage
Centre-right	25.7%
Centre-left	19.1%
Right	13%
Other	9%
Centre	7.2%
Centre-left to left	4.3%
Centre-right to right	3.2%
Right to left-right	4.9%
Centre to centre-right	1.7%
Centre to centre-left	1.7%

EXPLORAR O CONTIDO- frecuencias das palabras

DASHBOARD ParlaMint-ES-GA 4.0 (Galician parliament)

PARLAMINT-ES-GA 4.0 (GALICIAN PARLIAMENT) **CORPUS INFO**

- Concordance**
Examples of use in context
- Parallel Concordance**
Translation search
- Wordlist**
Frequency list
- Keywords**
Terminology extraction
- Text type analysis**
Statistics of the whole corpus
- Trends**
Diachronic analysis: neologisms

You are using NoSketch Engine. These tools are only available in Sketch Engine

WORDLIST ParlaMint-ES-GA 4.0 (Galician parliament)

BASIC **ADVANCED** **ABOUT**

the **words** **all**

- lemmas
- syntactic addr
- UD PoS tag
- UD features
- ID of token
- UD dependency label
- lemma of the dependency head

Exclude these words:

Include constants:

All

Frequency min: 5 Frequency max: 5

word forms:

- Single list
- Display as

display:

none (the whole corpus)

Text types

00

EXPLORAR O CONTIDO- tipos de busca

DASHBOARD

ParlaMint-ES-GA 4.0 (Galician parliament)

PARLAMINT-ES-GA 4.0 (GALICIAN PARLIAMENT)

CONCORDANCE
Examples of use in context

Wordlist
Frequency list

Trends
Diachronic analysis, neologisms

Corpus Info Manage Corpus

You are using NoSketch Engine. These tools are not available.

CONCORDANCE

ParlaMint-ES-GA 4.0 (Galician parliament)

BASIC ADVANCED ABOUT

Query type
simple
lemma
phrase
word
character
COL

Part of speech
any

Lemma
gasto
✓ A-a*

Scope
none (the whole corpus)

Match
none

Filter context

Text types

GO

EXPLORAR O CONTIDO- resultados da busca en contexto

CONCORDANCE

ParlaMint-ES-GA 4.0 (Galician parliament)

lemma **gasto** • 5,735

282.04 per million tokens • 0.028%



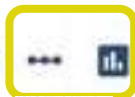
Download



View options



GD EX



Frequency

Collocations

KWIC



Details

Left context

KWIC

Right context

1	<input type="checkbox"/>	BalseiroJoséMan...) millóns de euros no ano 2021 , é un	gasto	disparado para dar soporte a todas as estrutura
			NOUN ADP NOUN ADP DET NOUN NUM PUNCT AUX DET	NOUN	VERB ADP VERB NOUN ADP DET DET NOUN
2	<input type="checkbox"/>	BalseiroJoséMan...	mente unha cotización que lles pode supoñer un	gasto	medio de máis de douscentos euros ao mes
			DET NOUN PRON PRON VERB VERB DET	NOUN	ADJ ADP ADV ADP NUM NOUN ADP DET NOUN
3	<input type="checkbox"/>	BalseiroJoséMan...	istentar unha cotización tan elevada , sufra un	gasto	extra de aproximadamente tres mil euros ao
			VERB DET NOUN ADV ACJ PUNCT VERB DET	NOUN	ADJ ADP ADV NOUN NOUN NOUN ADP DE
4	<input type="checkbox"/>	ÁlvarezLuisManu...	sabemos que o capítulo I se refire aos	gastos	globais de persoal , que inclúen , evider
			VERB SCONJ DET NOUN NOUN PRON VERB ACJ DET	NOUN	ADJ ADP NOUN PUNCT PRON VERB PUNCT
5	<input type="checkbox"/>	PradoPaula • 20...	vertiu a España de que tiña que controlar o	gasto	corrente porque o endebedamento excede — d
			ERS DET PROPIN ADP PRON VERB SCONJ VERB DET	NOUN	ADJ SCONJ DET NOUN VERB SYM VEI
6	<input type="checkbox"/>	DíazNoaSusana • ...	, apúntase un descenso moi importante no	gasto	das adquisicións de bibliotecas e na d
			NCT VERB PRON DET NOUN ADV ADJ ADP DET	NOUN	ADP DET NOUN ADP NOUN CONJ ADP DET
7	<input type="checkbox"/>	VázquezSandra • ...	atorio da Cultura Galega que amosa que o	gasto	en cultura e lingua orzado pola Administraci
			N ADP DET NOUN ADJ PRON VERB SCONJ DET	NOUN	ADP NOUN CONJ NOUN VERB ADP DET NOUN
8	<input type="checkbox"/>	NóvoaMarta • 20...	je * Bruselas insta a España a cuidar el	gasto	público y le reprocha los retrasos con los
			NU PUNCT PROPIN VERB DET PROPIN ADP VERB PRON	NOUN	ADJ NOUN VERB VERB NOUN NOUN ADP NOUN

EXPLORAR O CONTIDO- análise dos resultados

CONCORDANCE

lemma **gasto** • 8736

3034 permutations • 1.00%

Details

	Left context	KWIC	Right context
BabeiroJoséMan...	lémbras de euros no ano 2021	gasto	disparado para dar soporte
BabeiroJoséMan...	mente esta cotización que les pode supoñer un	gasto	medio de máis de dousos
BabeiroJoséMan...	dentro unha cotización tan elevada	gasto	extra de aproximadamente
AlvarezLuisMan...	abonos que o capítulo 1 se refire aos	gastos	globais de persoal
PardoPablo • 20...	verte a España de que máis que controla o	gasto	corres
DíazNúñezGuad...	apuntase un deceno no importante	gasto	das

FREQUENCY

BASIC ADVANCED ABOUT

Select an attribute and its position in the concordance

UD PoS tag

Left context

Right context

KWIC

More presets

- KWIC WORD FORMS
- KWIC TAGS
- KWIC LEMMAS
- TEXT TYPES
- LINE DETAILS

Frequency

CHANGE CRITERIA BACK TO CONCORDANCE

Show relative frequency

UD PoS tag

ADJ

ADP

FREQUENCY

BASIC ADVANCED ABOUT

Select an attribute and its position in the concordance

lemma (lowercase)

Left context

Right context

KWIC

Group by first token

FTC (Items: 1.000 total frequency)

Lemma (lowercase)

1	<input type="checkbox"/>	público
2	<input type="checkbox"/>	social
3	<input type="checkbox"/>	corrente
4	<input type="checkbox"/>	sanitario
5	<input type="checkbox"/>	farmacéutico
6	<input type="checkbox"/>	financeiro
7	<input type="checkbox"/>	total

8	<input type="checkbox"/>	medio
9	<input type="checkbox"/>	extraordinario
10	<input type="checkbox"/>	real
11	<input type="checkbox"/>	privado
12	<input type="checkbox"/>	educativo
13	<input type="checkbox"/>	produtivo
14	<input type="checkbox"/>	electoral

EXPLORAR O CONTIDO- combinaciones de palabras

CONCORDANCE Palabras ES-GA 4.0 (palabras)

letras **gato** = 8,735
28.14 por millón de letras = 0.02%

--- **COLLOCATIONS**

BASIC ADVANCED ABOUT

Alfabeto [↑] Range [↑]

UD PoS tag -5 -4 -3 -2 -1 KWIC 1 2 3 4 5

Custom range

Collocations CHANGE CRITERIA BACK TO CONCORDANCE

UD PoS tag	Concordances ¹	Distances	T score
<input type="checkbox"/> NUMSYM	101	17,450	11.85
<input type="checkbox"/> ADV	2,827	1,285,799	46.45
<input type="checkbox"/> ADPDET	2,407	1,504,983	42.25
<input type="checkbox"/> VERBPRECONJPHON	28	9,830	4.77
<input type="checkbox"/> CONJ	1,143	735,967	27.75
<input type="checkbox"/> NUM	384	259,180	15.91
<input type="checkbox"/> ADP	3,953	3,075,100	48.29
<input type="checkbox"/> SYM	109	140,558	10.58
<input type="checkbox"/> PUNCT	2,995	2,456,643	41.86
<input type="checkbox"/> ADV	1,218	1,193,418	25.95
<input type="checkbox"/> DET	3,627	3,329,973	43.77
<input type="checkbox"/> NOUN	4,443	4,109,588	48.90
<input checked="" type="checkbox"/> VERB	3,529	3,483,140	38.52

FREQUENCY

BASIC ADVANCED ABOUT

Return an attribute and its position in the concordance [↑]

lemma (lowercase)

left context

COLLOCATION 1 (HIGHLIGHTED)

991 items, 2,958 total frequencies

Lemma (lowercase)	
<input type="checkbox"/> decir	1
<input type="checkbox"/> medrar	9
<input type="checkbox"/> permitir	18
<input type="checkbox"/> aumentar	11
<input type="checkbox"/> asumir	12
<input type="checkbox"/> financiar	13
<input type="checkbox"/> reforzar	14
<input type="checkbox"/> recortar	15
<input type="checkbox"/> ter	1
<input type="checkbox"/> incrementar	7
<input type="checkbox"/> reducir	5
<input type="checkbox"/> haber	4
<input type="checkbox"/> poder	8
<input type="checkbox"/> hacer	6
<input type="checkbox"/> ir	7

EXPLORAR O CONTIDO- concordancias paralelas

simple **nation** • 204,781
136.61 per million tokens • 0.014%

Collocations

Lemma (lowercase) C

- united
- unite
- turkish
- beloved
- convention
- nation
- entire
- devolve
- great
- hungarian

PARALLEL CONCORDANCE

ParlaMint-XX-en-4.0 (European parliaments, translation to English)

Sample 1 (left) • 204,781
136.61 per million tokens • 0.014%

Sample 2 (right) • 250
8.17 per million tokens • 1.00017%

UA-en • 2023-09-05
Convinced that the restoration of Ukraine will be a matter of honor for all **nations**, and countries that share the principles for which Ukrainians and Ukrainians today lay down their lives.

TR-es • 2022-11-16
The Turkish **nation** is a fortunate nation in this regard.

ES-es • 2022-11-03
Nations, international judicial meeting, and Victoria Rosell, both victims of different versions of judicial war against her.

Переконані, що відновлення України стане справою честі усіх народів та країн, які поділяють принципи, за які українці і українки сьогодні віддають свої життя.

Türk milletini bu bakımdan şanslı bir milletdir.

También se justificaron con los votos para ingresar en el Consejo General del Poder Judicial a José Ricardo de Prada, el único juez español en el plantel de jueces internacionales de Naciones Unidas, y a Victoria Rosell, ambas víctimas de diferentes versiones de guerra judicial en su contra.

Speech party_gramadan	Frequency	Rel (%)	Per-million tokens
Right to far-right	1,320	546.14	0.88
Centre-right	1,001	120.00	0.67
Right	824	250.59	0.55
Centre-left	602	71.32	0.46
.	160	33.41	0.11
Centre-right to right	74	31.21	0.06
Centre	60	25.31	0.04
Centre to centre-right	48	23.31	0.03
Far-right	40	44.84	0.03
Centre-left to left	46	20.79	0.03



UA-en • 2023-09-05

In short, we survived as a **nation** and a **nation**.

Одним словом, ми виживали як держава як нація.

/We survived as a state and a nation./

Casos de uso

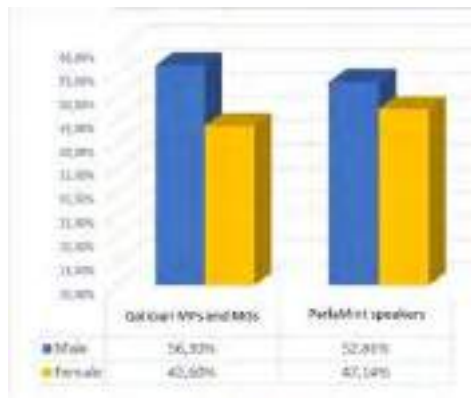


Xénero e representación no Parlamento de Galicia

1. Unha presenza relativamente equilibrada de homes e mulleres no Parlamento, correspóndese cun acceso equitativo ao uso da palabra?
2. Que se pode deducir respecto á representación parlamentaria analizando a intersección entre o número de intervencións, o número de mencións, o xénero e a identidade dos oradores que participan nos debates do Parlamento galego?

Os resultados respaldan estudos previos:

- o relativo equilibrio de xénero na representación parlamentaria non garante necesariamente unha participación igualitaria;
- disparidades na frecuencia das intervencións e as referencias, moi relacionadas coa posición política dos oradores.



Gender	AMs	AM (%)	PMs	PM (%)
Male	25,485	56.6%	31,196	69.2%
Female	19,566	43.4%	13,855	30.8%

Table 4: Active and passive mentions by gender in number of mentions and percentage of the total number of mentions

Role	Speeches	Words	% of Total Speeches	% of Total Words
Male Chair	37,885	799k	45.6%	4.5%
Female Chair	7,103	141k	8.5%	0.8%
Male Regular	20,735	9045k	24.9%	51.1%
Female Regular	17,305	7601k	20.8%	43.1%
Male Guest	28	20k	0.03%	0.1%
Female Guest	22	28k	0.03%	0.1%

Table 1: Distribution of speeches and words according to gender and role in the dataset.

Redes de poder

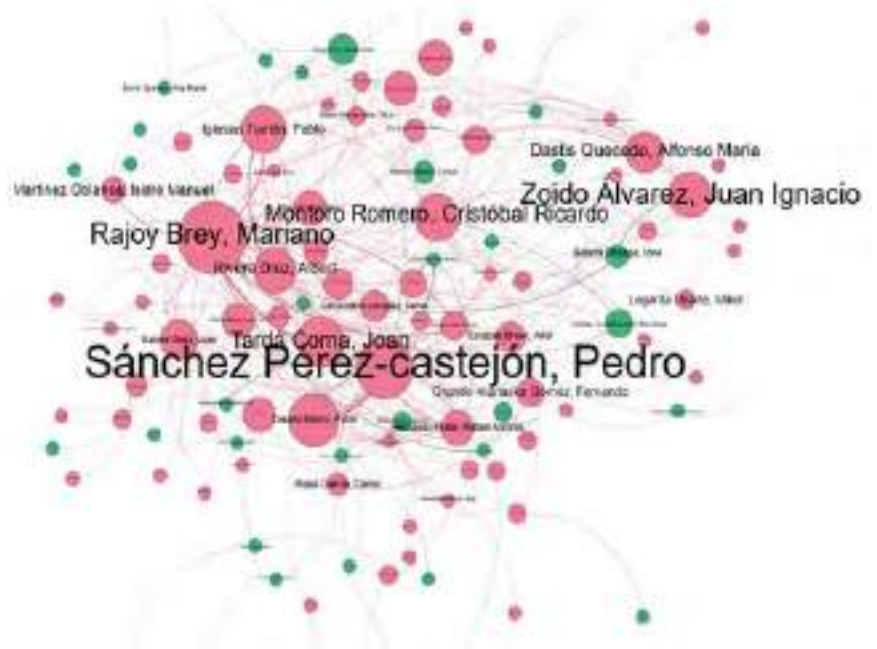
Correspóndense as redes que xorden analizando as mencións entre deputados coas estruturas de poder?

Si, reflicten o poder estrutural existente (por exemplo, os cargos ministeriais) dos deputados dentro de temas específicos.

Angermeier et al. (2022)

ParlaMint. Networks of Power

<https://www.clarin.eu/impact-stories/networks-power-gender-analysis-european-parliaments>

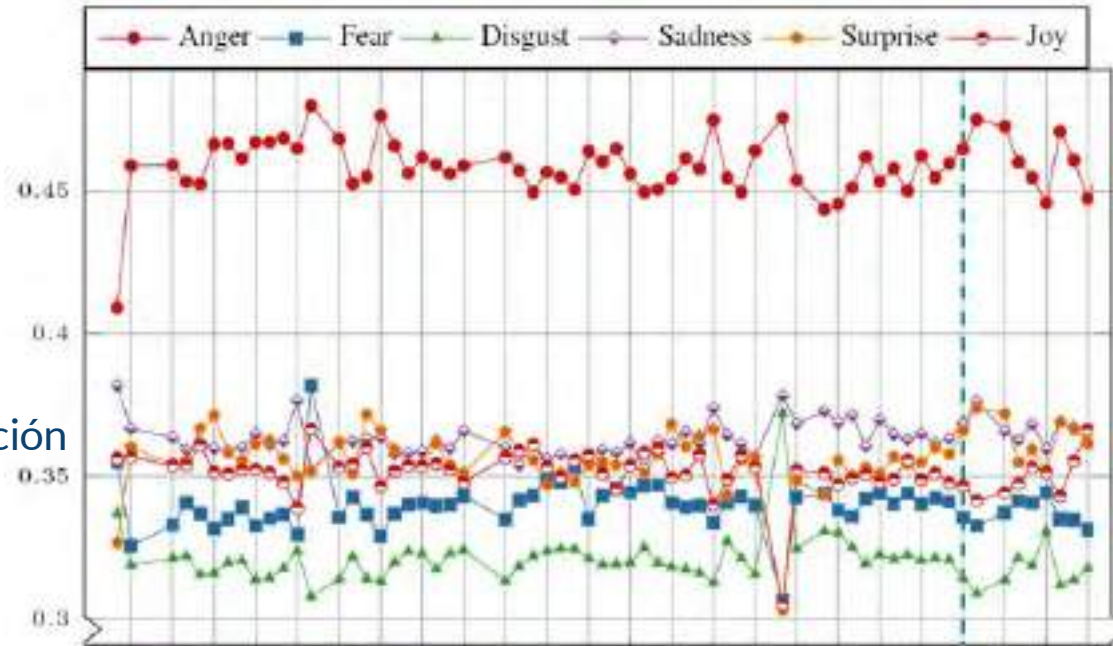


Análise de emocións

No Parlamento turco :

- a análise das emocións reflicte a polarización da escena política;
- domina a ira;
- o partido que está ao poder mostra emocións máis estables que a oposición

Kurtoğlu and Çöltekin (2022).
Emotions Running High?

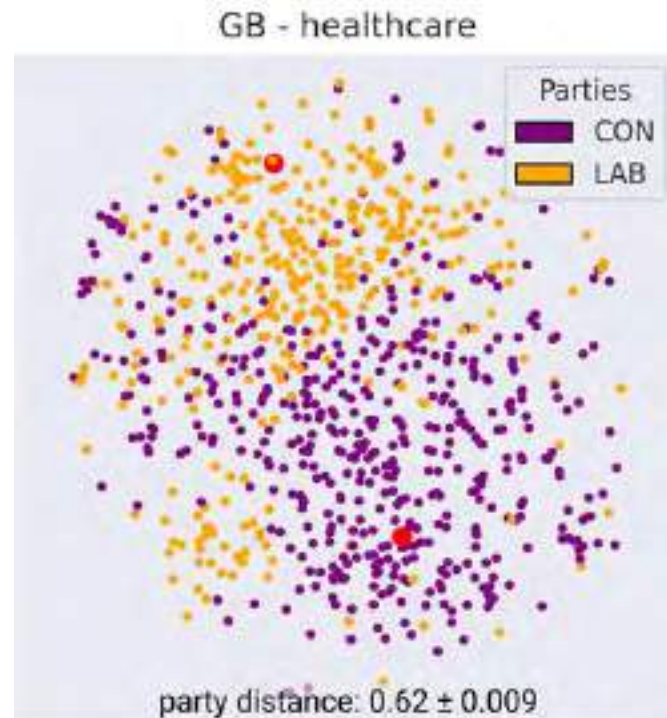


Polarización política

Como polarizan os parlamentos (GB, HU, UA, SI) determinados temas ao longo do tempo?

1. Uso de LLM multilingües para análise de discurso e sentimentos.
2. Cálculo e visualización de *embeddings* SBERT para detectar diferenzas no discurso dos partidos políticos.

Kryvenko et al. (2023)
Splitting Lips. Polarization through
Parliamentary Speech



Ligazóns útiles



Publicacións

- Erjavec T. et al. (2024). [ParlaMint II: Advancing Comparable Parliamentary Corpora Across Europe](#). *Language Resources and Evaluation* (preprint).
- Erjavec T. et al. (2023). [The ParlaMint corpora of parliamentary proceedings](#). *Language Resources and Evaluation* 57:415–448.

[Publicacións e presentacións do proxecto ParlaMint](#)



Corpus

Os corpus publicados por ParlaMint pódense descargar con licenza libre (CC BY):

- Tomaž Erjavec et al. (2023) *Multilingual comparable corpora of parliamentary debates ParlaMint 4.0*. <http://hdl.handle.net/11356/1859>
- Tomaž Erjavec et al. (2023) *Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 4.0*. <http://hdl.handle.net/11356/1860>
- Taja Kuzman et al. (2023) *Linguistically annotated multilingual comparable corpora of parliamentary debates in English ParlaMint-en.ana 4.0*. <http://hdl.handle.net/11356/1864>
- As mostras dos diferentes corpus, esquemas XML e scripts de validación pódense descargar do repositorio [GitHub](#) do proxecto.



Moitas grazas!

<https://www.clarin.eu/parlamint>

adina.vladu@usc.gal