Applying new technologies to give a step forward in terminology quality assessment and control
The territory of Basque language is inhabited by 3 150 000 people

Spanish and Basque are co-official languages
2 200 000 inhabitants

French is the only official language
300 000 inhabitants

Spanish and Basque are co-official only in one part of the territory
650 000 inhabitants
1. It is an agglutinating language

   • In Basque there are 15 base declination cases that can take the form of 22 variants.

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Spanish</th>
<th>Basque</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>2 forms</td>
<td>2 forms</td>
<td>275 forms</td>
</tr>
<tr>
<td>Adjective</td>
<td>1 form</td>
<td>4 forms</td>
<td>825 forms</td>
</tr>
</tbody>
</table>

   • Basque makes no gender distinction.

2. Both Basque vocabulary and terminology are undergoing a normalization process
UZEI, the Basque Centre for Terminology and Lexicography

- Non-profit organization
- Declared of public interest
- Working areas:

  - Terminology
  - Lexicography
  - Language technologies
  - Translation
  - Standardization
  - Modernization
The Basque Government is the chief agent in the language planning and the terminological standardization of Basque.
Applying new technologies to give a step forward in terminology quality assessment and control
• SASK06 (Saiotek 2006)
• IDITE (IKT GF 2006)
• TEIS07 (Saiotek 2007)
• GAIKA (IKT 2007)
• Hizperri (Saiotek 2008)
• Corpusa2008 (HPS-IKT 2008)
• Hizperri (Saiotek 2009)
• euLEXpro (HPS IKT 2009)
• IMEBI (e-GFA 2009)
• iTXEK - (Saiotek -2010)
• GAIMET (Saiotek 2011)
• Semtrack (Gaitek 2012-2013)
• Gaimet (e-GFA 2012)
• Sumatu (Saiotek 2013)
• LEX2 (e-GFA 2013)
• OLITE (Gaitek 2013-2015)
• IDITEsuite (GFA IKT 2015 innovación)

Spelling checker (Microsoft, LibreOffice)

Public Bank of Translation Memories

Parallel lexicon extraction
What is IDITE?

How does IDITE work?

IDITE and terminology evaluation and control

Who uses IDITE?

Possible future applications
What is IDITE?

- Lexical database
- Terminology
- General lexicon
- Computer application
- Collection of linguistic rules
Constituents of IDITE’s lexical database:

1. UTH, the terminological part which is tagged for NLP and comprises 94 400 terminological records which are classified by subject.

2. euLEX, the vocabulary part which comprises 145 600 entries.

They include up-to-date information on forms considered both as correct and incorrect by the linguistic authorities in Basque terminology and lexicography, respectively.
The rejected, not preferred and unknown forms are listed

<table>
<thead>
<tr>
<th>IDITE bostena</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>eritzi (eritzi)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erabili</td>
<td>iritzi</td>
<td>Hiztegi orokorra</td>
</tr>
<tr>
<td><strong>domeka (domeka)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hobe</td>
<td>igande</td>
<td>Hiztegi orokorra</td>
</tr>
<tr>
<td><strong>Japon (Japon)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erabili</td>
<td>Japonia</td>
<td>Hiztegi orokorra</td>
</tr>
<tr>
<td><strong>agroturismo (agroturismo)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hobe</td>
<td>nekazaritza-turismoko</td>
<td>Merkataritza</td>
</tr>
<tr>
<td></td>
<td>establezimendu</td>
<td></td>
</tr>
<tr>
<td>Hobe</td>
<td>nekazaritza-turismo</td>
<td>Merkataritza</td>
</tr>
<tr>
<td><strong>adin txikiko (adin txikiko)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erabili</td>
<td>adingabe</td>
<td>Hiztegi orokorra</td>
</tr>
<tr>
<td>Erabili</td>
<td>adingabe/adingabeko</td>
<td>Zuzentidea</td>
</tr>
</tbody>
</table>
The rejected, not preferred and unknown forms are listed.
The rejected, not preferred and unknown forms are listed

<table>
<thead>
<tr>
<th>IDITE txostena</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>eritzi (eritzi)</td>
<td></td>
</tr>
<tr>
<td>Erabili</td>
<td>iritzi</td>
</tr>
<tr>
<td>domeka (domeka)</td>
<td>Dialectal variants</td>
</tr>
<tr>
<td>Hobe</td>
<td>igande</td>
</tr>
<tr>
<td>Japon (Japon)</td>
<td></td>
</tr>
<tr>
<td>Erabili</td>
<td>Japonia</td>
</tr>
<tr>
<td>agroturismo (agroturismo)</td>
<td></td>
</tr>
<tr>
<td>Hobe</td>
<td>nekazaritza-turismoko</td>
</tr>
<tr>
<td></td>
<td>establezimendu</td>
</tr>
<tr>
<td>Hobe</td>
<td>nekazaritza-turismo</td>
</tr>
<tr>
<td>adin txikiko (adin txikiko)</td>
<td></td>
</tr>
<tr>
<td>Erabili</td>
<td>adingabe</td>
</tr>
<tr>
<td>Erabili</td>
<td>adingabe/adingabeko</td>
</tr>
</tbody>
</table>
The rejected, not preferred and unknown forms are listed

<table>
<thead>
<tr>
<th>General vocabulary</th>
<th>Prescribed geographical names</th>
</tr>
</thead>
<tbody>
<tr>
<td>eritz[i] (eritzi)</td>
<td></td>
</tr>
<tr>
<td>Erabili</td>
<td>iritzi</td>
</tr>
<tr>
<td>Hiztegi orokorra</td>
<td></td>
</tr>
<tr>
<td>domeka (domeka)</td>
<td></td>
</tr>
<tr>
<td>Hobe</td>
<td>igande</td>
</tr>
<tr>
<td>Hiztegi orokorra</td>
<td></td>
</tr>
<tr>
<td>Japon (Japon)</td>
<td></td>
</tr>
<tr>
<td>Erabili</td>
<td>Japonia</td>
</tr>
<tr>
<td>Hiztegi orokorra</td>
<td></td>
</tr>
<tr>
<td>agroturismo (agroturismo)</td>
<td></td>
</tr>
<tr>
<td>Hobe</td>
<td>nekazaritza-turismoko</td>
</tr>
<tr>
<td>Merkatatitza</td>
<td></td>
</tr>
<tr>
<td>establezimendu</td>
<td></td>
</tr>
<tr>
<td>Hobe</td>
<td>nekazaritza-turismo</td>
</tr>
<tr>
<td>Merkatatitza</td>
<td></td>
</tr>
<tr>
<td>adin txikiko (adin txikiko)</td>
<td></td>
</tr>
<tr>
<td>Erabili</td>
<td>adingabe</td>
</tr>
<tr>
<td>Hiztegi orokorra</td>
<td></td>
</tr>
<tr>
<td>Erabili</td>
<td>adingabe/adingabeko</td>
</tr>
<tr>
<td>Zuzentidea</td>
<td></td>
</tr>
</tbody>
</table>
The rejected, not preferred and unknown forms are listed

Terminological variants recommended by BGTAC*

* BGTAC: Basque Government’s Terminology Assessment Commission
The rejected, not preferred and unknown forms are listed

<table>
<thead>
<tr>
<th>IDITE txostena</th>
<th>Ezezagunak</th>
</tr>
</thead>
<tbody>
<tr>
<td>eritzi (eritzi)</td>
<td>iritzi</td>
</tr>
<tr>
<td>Erabili</td>
<td>Hiztegi orokorra</td>
</tr>
<tr>
<td>domeka (domeka)</td>
<td>igande</td>
</tr>
<tr>
<td>Hobe</td>
<td>Hiztegi orokorra</td>
</tr>
<tr>
<td>Japon (Japon)</td>
<td>Japonia</td>
</tr>
<tr>
<td>Erabili</td>
<td>Hiztegi orokorra</td>
</tr>
<tr>
<td>agroturismo (agroturismo)</td>
<td>nekazaritza-turismo</td>
</tr>
<tr>
<td>Hobe</td>
<td>Merkataritza</td>
</tr>
<tr>
<td></td>
<td>establezimendu</td>
</tr>
<tr>
<td>Hobe</td>
<td>Merkataritza</td>
</tr>
<tr>
<td>adin txikiko (adin txikiko)</td>
<td>adingabe</td>
</tr>
<tr>
<td>Erabili</td>
<td>Hiztegi orokorra</td>
</tr>
<tr>
<td></td>
<td>(General vocabulary)</td>
</tr>
<tr>
<td>Erabili</td>
<td>adingabe/adingabeko</td>
</tr>
<tr>
<td></td>
<td>Zuzentzieta</td>
</tr>
<tr>
<td></td>
<td>(Law)</td>
</tr>
</tbody>
</table>

* BGTAC: Basque Government’s Terminology Assessment Commission
IDITE can be used for the evaluation and control of vocabulary and terminology used in:

1. Texts that have been recently written. This enables:
   - Ensuring the good quality of vocabulary and terminology.
   - Detecting and correcting vocabulary, terminology and spelling mistakes in a quick and economical way.

2. Text written at different times. This provides a picture of the diachronic changes undergone by terminology from a particular field.
● IDITE fulfills the function of conventional spelling checkers.

● It can be used for **neologism detection**, as neologisms may be present among unknown forms listed in IDITE’s report.

UZEI checks the texts of the main Basque newspaper using IDITE and systematically incorporates in Euskalterm its terminology proposals for the neologisms detected.
UZEI integrates in Euskalterm terminology proposals for the neologisms detected using IDITE

We have found 1 terms. Click icon if you want to open and see the whole Terminology Record.

nomophobia

- Psikologia > Nortasuna -

**eu** nomofobia

**Definizioa:**
Teleforno mugikorik eskura ez izateak sortuko antsietate edo egonezina.

**es** nomofobia  
**fr** nomophobie  
**en** nomophobia

[Kontsulta-fitsak] [2016]
IDITE offers the option to create personal dictionaries.

When including a new form in IDITE’s personal dictionary, the user can indicate:

- If it is a preferred, correct but not preferred or incorrect form.
- If it belongs to general vocabulary or, if it is a term, which thematic domain it belongs to.

Correction and improvement proposals based on the information contained in personal dictionaries are made by the tool in the same way shown in the example report.
<table>
<thead>
<tr>
<th>Who uses IDITE?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Provincial Council of Biscay</strong></td>
</tr>
<tr>
<td><strong>Provincial Council of Gipuzkoa (and its capital’s San Sebastian’s - city hall)</strong></td>
</tr>
<tr>
<td><strong>Basque Parliament</strong></td>
</tr>
<tr>
<td><strong>HABE, the institute for the basquization and the literacy training of adults</strong></td>
</tr>
<tr>
<td><strong>Euskaltzaindia, Basque Language Royal Academy</strong></td>
</tr>
</tbody>
</table>

And, of course... UZEI
The technology at the base of IDITE could be applied to create new versions of IDITE for languages other than Basque.

UZEI and the University of Salamanca are nowadays studying the possibility to tailor the technology of IDITE to develop a Spanish version for detecting neologisms in texts with a high specialization level.
IDITE is the only tool available for automatically checking the adequacy of vocabulary and terminology contained in texts written in Basque.

It has been fully developed using UZEI’s technology and linguistic resources.

It can also be used for neologism detection and for analyzing the diachronic changes undergone by terminology.

IDITE’s Spanish version will be developed in the near future.
For further information:

www.uzei.eus

barrate@uzei.eus