# Biomedical Abbreviation Recognition and Resolution 2nd Edition (BARR2)

**IberEval 2018 | SEPLN 2018**

18 September 2018. Seville, Spain

**Plan TL**
Plan de Impulso de las
Tecnologías del Lenguaje

- Detecting & resolving abbreviations is critical for almost any NLP task: machine translation, IR, IE, NER, text categorization, summarization, linguistic annotation,...

- Abbreviation resolution approaches are highly portable/adaptable across different languages

- Abbreviation recognition/resolution is key both across languages and domains

- For clinical & biomedical NLP resolution of abbreviation is a critical initial step

- Abbreviations the most challenging aspect for Spanish medical translators

- The BARR2 tasks requires for *clinical case studies*:

  Task 1: Detect explicit occurrences of abbreviation-definition 8short form-long form) pairs

  **Task 2: Detect and resolve all abbreviations mentions returning their definition**

  More info, dates and registration: http://temu.bsc.es/BARR2

Mujer de 84 años sin **ACM**. Niega hábitos tóxicos. Parcialmente dependiente **ABVD**. Vive en residencia. Antecedentes de **HTA**, **DLF** y **FA** antiagregada. Ictus **POCI ACP** izquierda en 2008, etiología cardioembólica.

Mujer de 84 años sin **alergias medicamentosas conocidas**. Niega hábitos tóxicos. Parcialmente dependiente **adriamycin bleomycin vinblastine and dacarbazin**. Vive en residencia. Antecedentes de **hipertensión arterial**, **depresión a largo plazo** y **fibrilación auricular** antiagregada. **Ictus circulación posterior, arteria cerebral posterior** izquierda en 2008, etiología cardioembólica.

Plan TL
Plan de Impulso de las
Tecnologías del Lenguaje

Varón de 58 años de edad en el momento del trasplante, el 5 de octubre de 1998.

En programa de diálisis peritoneal desde enero del mismo año por IRC progresiva secundaria a glomerulonefritis tipo IgA con esclerosis focal y segmentaria asociada diagnosticada en 1984.

Otros antecedentes de interés: hipertensión arterial diagnosticada en 1982, en tratamiento farmacológico, cardiopatía isquémica por enfermedad coronaria con angioplastia transluminal de la coronaria derecha en 1994 (desde entonces asintomático, pruebas de esfuerzo posteriores negativas), dislipemia tipo II. Intervenido de criptorquídia bilateral y amigdalectomizado en juventud.
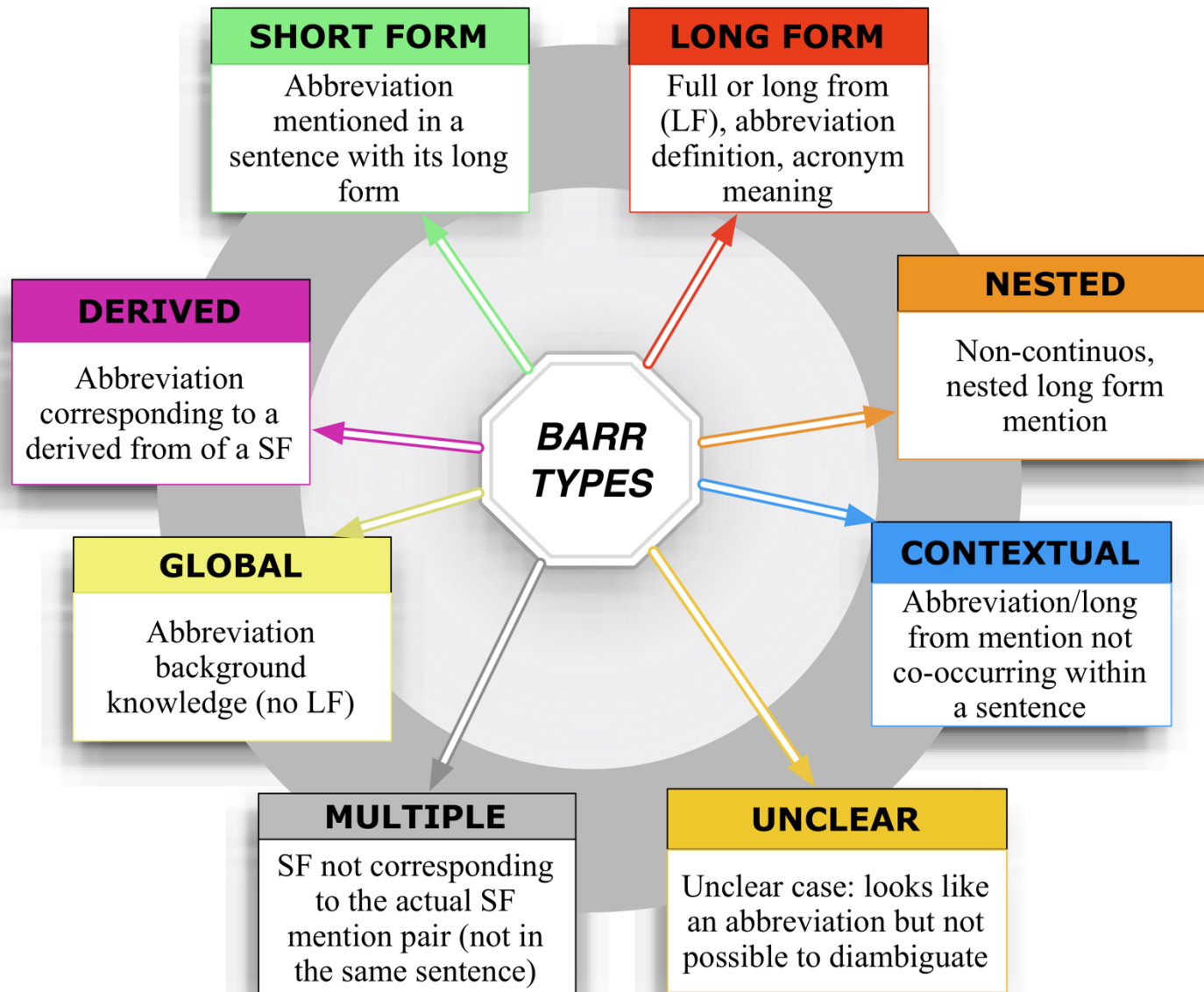
El trasplante renal cursó sin incidencias (riñón izquierdo en fosa ilíaca derecha con 2 arterias en parche único y vena que se abocan a ilíacas externas), con función eficaz inmediata del injerto, sin rechazo agudo y niveles elevados de ciclosporina sin toxicidad clínica aparente. El paciente recibe tratamiento con ciclosporina 8mg/Kg, micofenolato mofetilo 2gr/día, y prednisona 1mg/Kg, disminuyendo progresivamente las cifras en controles posteriores.

En septiembre de 2000 (a los 2 años del trasplante), el paciente refiere por primera vez sensación de inestabilidad a la marcha y temblores en las manos que aumentan progresivamente iniciándose estudio por Servicio de neurología. En el momento del estudio el paciente presenta inestabilidad de la marcha, parestesias en ambas extremidades inferiores (EEII) y temblores en manos. A la exploración física no se observa déficit motor pero sí disminución de la sensibilidad vibratoria en ambas EEII, reflejos osteotendinosos (ROT) vivos y reflejo cutaneoplantar (RCP) en extensión. La resonancia magnética nuclear (RMN) no mostró lesiones a nivel cerebral o espinal, el análisis del líquido cefalorraquídeo (LCR) mostró pleocitosis linfocitaria sin bandas oligoclonales, los potenciales evocados resultaron alterados a nivel tibial y fueron los test serológicos, ELISA, y la reacción en cadena de la polimerasa (PCR) los que fueron positivos para HTLV-I en LCR y sangre en el primer caso y en sangre la PCR. Se demostró una alta carga viral.

El paciente fue diagnosticado de paraparesia espástica tropical (TSP), y a pesar de repetidos pulses de 6-Metil-Prednisolona, evoluciona de forma progresiva. Ha rechazado tratamiento con Interferón por la posibilidad de rechazo del trasplante.Desde punto vista de la función renal, en todo momento se ha mantenido correcta con niveles de creatinina oscilantes entre 1-12.

Ha sid[...]
mediar[...]
cistoli[...]

En est[...]
ruedas[...]

| # Document_ID | StartOffset | EndOffset | Abbreviation | Definition | Definition_lemmatized |
|---|---|---|---|---|---|
| S0004-06142005001000011-1 | 1034 | 1036 | Kg | kilogramo | kilogramo |
| S0004-06142005001000011-1 | 1031 | 1033 | mg | miligramo | miligramo |
| S0004-06142005001000011-1 | 196 | 199 | IgA | inmunoglobulina a | inmunoglobulina a |
| S0004-06142005001000011-1 | 2057 | 2060 | LCR | líquido cefalorraquídeo | líquido cefalorraquídeo |
| S0004-06142005001000011-1 | 1594 | 1598 | EEII | extremidades inferiores | extremidad inferior |
| S0004-06142005001000011-1 | 1009 | 1011 | gr | gramo | gramo |
| S0004-06142005001000011-1 | 982 | 984 | Kg | kilogramo | kilogramo |
| S0004-06142005001000011-1 | 979 | 981 | mg | miligramo | miligramo |
| S0004-06142005001000011-1 | 1963 | 1968 | ELISA | enzyme-linked immunosorbent assay | enzyme-l[...] |
| S0004-06142005001000011-1 | 2207 | 2210 | TSP | tropical spastic paraparesis | tropical spastic[...] |
| S0004-06142005001000011-1 | 2012 | 2015 | PCR | reacción en cadena de la polimerasa | reacción[...] |
| S0004-06142005001000011-1 | 1808 | 1811 | LCR | líquido cefalorraquídeo | líquido cefalorraquídeo |
| S0004-06142005001000011-1 | 1663 | 1666 | RCP | reflejo cutaneoplantar | reflejo cutaneoplantar |
| S0004-06142005001000011-1 | 1626 | 1629 | ROT | reflejos osteotendinosos | reflejo osteoten[...] |

# Abbreviation mention types

**SHORT FORM**
Abbreviation mentioned in a sentence with its long form

**LONG FORM**
Full or long from (LF), abbreviation definition, acronym meaning

**DERIVED**
Abbreviation corresponding to a derived from of a SF

**NESTED**
Non-continuos, nested long form mention

*BARR TYPES*

**GLOBAL**
Abbreviation background knowledge (no LF)

**CONTEXTUAL**
Abbreviation/long from mention not co-occurring within a sentence

**MULTIPLE**
SF not corresponding to the actual SF mention pair (not in the same sentence)

**UNCLEAR**
Unclear case: looks like an abbreviation but not possible to diambiguate

# Biomedical Abbreviation Recognition and Resolution 2nd Edition (BARR2)

IberEval 2018 | SEPLN 2018

18 September 2018. Seville, Spain

**Plan TL**
Plan de Impulso de las
Tecnologías del Lenguaje

| Tentative dates | |
|---|---|
| **20th April 2018** | **Release of sample data (sub-track 1 and sub-track 2)** |
| **16th May 2018** | Training corpus available (sub-track 1 and sub-track 2) |
| **23rd May 2018** | Development corpus available (sub-track 1 and sub-track 2) |
| **25th May 2018** | Test corpus available (sub-track 1 and sub-track 2) |
| **7th May 2018** | Submission of the results (sub-track 1 and sub-track 2) |
| **10th May 2018** | Publication of results (sub-track 1 and sub-track 2) |
| **11th June 2018** | Working notes submission |
| **1st July 2018** | Release of the working notes reviews |
| **15th July 2018** | Camera ready paper submission |
| **18 September 2018** | IberEval 2018 Workshop |

More info, dates and registration: http://temu.bsc.es/BARR2

Plan TL
Plan de Impulso de las

**IBEREVAL'17**  **TASKS**  **SCHEDULE**

IberEval 2017

Evaluation of Human Language Technologies for Iberian languages

## Tasks

We are pleased to announce the following tasks in IberEval 2017:

- Biomedical Abbreviation Recognition and Resolution
- Classification Of Spanish Election Tweets (COSET)
- Collective Elaboration of a Coreference Annotated Corpus for Portuguese Texts
- Multilingual Web Person Name Disambiguation (M-WePNaD)
- Stance and Gender detection in tweets on Catalan Independence

### The Biomedical Abbreviation Recognition and Resolution (BARR) track: benchmarking, evaluation and importance of abbreviation recognition systems applied to Spanish biomedical abstracts

Ander Intxaurrondo[1], Martin Pérez-Pérez[2], Gael Pérez-Rodríguez[2], Jose Antonio López-Martín[3], Jesus Santamaría[1], Santiago de la Peña[1], Marta Villegas[1,4], Saber Ahmad Akhondi[5], Alfonso Valencia[4], Analia Loureno[2], Martin Krallinger[1*]

[1]Centro Nacional de Investigaciones Oncológicas (CNIO),
[2]ESEI - Department of Computer Science, University of Vigo,
[3]Hospital 12 de Octubre - Madrid,
[4]Barcelona Supercomputing Center (BSC),
[5]Elsevier Content & Innovation

**Abstract.** Healthcare professionals are generating a substantial volume of clinical data in narrative form. As healthcare providers are confronted with serious time constraints, they frequently use telegraphic phrases, domain-specific abbreviations and shorthand notes. Efficient clinical text processing tools need to cope with the recognition and resolution of abbreviations, a task that has been extensively studied for English documents. Despite the outstanding number of clinical documents written worldwide in Spanish, only a marginal amount of studies has been published on this subject. In clinical texts, as opposed to the medical literature, abbreviations are generally used without their definitions or expanded forms. The aim of the first Biomedical Abbreviation Recognition and Resolution (BARR) track, posed at the IberEval 2017 evaluation campaign, was to assess and promote the development of systems for generating a sense inventory of medical abbreviations. The BARR track required the detection of mentions of abbreviations or short forms and their corresponding long forms or definitions from Spanish medical abstracts. For this track, the organizers provided the *BARR medical document collection*, the *BARR corpus* of manually annotated abstracts labeled by domain experts and the *BARR-Markyt* evaluation platform. A total of 7 teams submitted 25 runs for the two BARR subtasks: (a) the identification of mentions of abbreviations and their definitions and (b) the correct detection of short form-long form pairs. Here we describe the BARR track setting, the obtained results and the methodologies used by participating systems. The BARR task summary, corpus, resources and evaluation tool for testing systems beyond this campaign are available at: http://temu.inab.org

## 1  Introduction

There is an increasing adoption of Electronic health records (EHRs) in the European Union, promoted both by national plans as well as European initiatives like the 2 billion Euro publicprivate partnership Innovative Medicines Initiative [18]. Electronic