

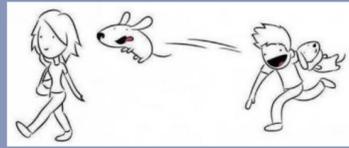
Multiword Expressions in different Spanish dialects

Diana Bogantes, Eric Rodriguez, Alejandro Arauco, Alejandro Rodriguez
 Université François-Rabelais Tours, Blois, France



Multiword Expressions

- ▶ Composed of two or more words
 - ▶ Meaning not derived from components (non-compositionality), i.e. not a literal meaning, e.g. *Spill the beans, kick the bucket*.
- In Spanish: *Echar los perros* = to throw the dogs → meaning not literal.

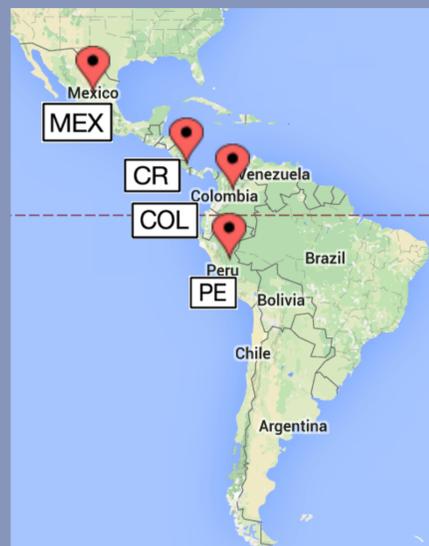


Challenges with Spanish MWEs

- ▶ Inflections (on gender, number, tense, mood, ...):
 - ▷ *Hacerse el loco* [to act crazy] → pretend not to see something
 - ▷ *me hice el loco, se hace la loca, se hicieron los locos, me hago la loca*
- ▶ Possibility of adding new words:
 - ▷ *Hablar paja* [have straw out of your mouth] → hablar paja
 - ▷ *Hablar pura paja*
- ▶ Possibility of substituting words:
 - ▷ *Tomar la posta* [to grab the relay stick] → to take the lead
 - ▷ *Tomar la batuta*
- ▶ Possibility of omitting base words:
 - ▷ *Se armó la gorda* [a fat woman was assembled] → a big problem started
 - ▷ *Se armó*

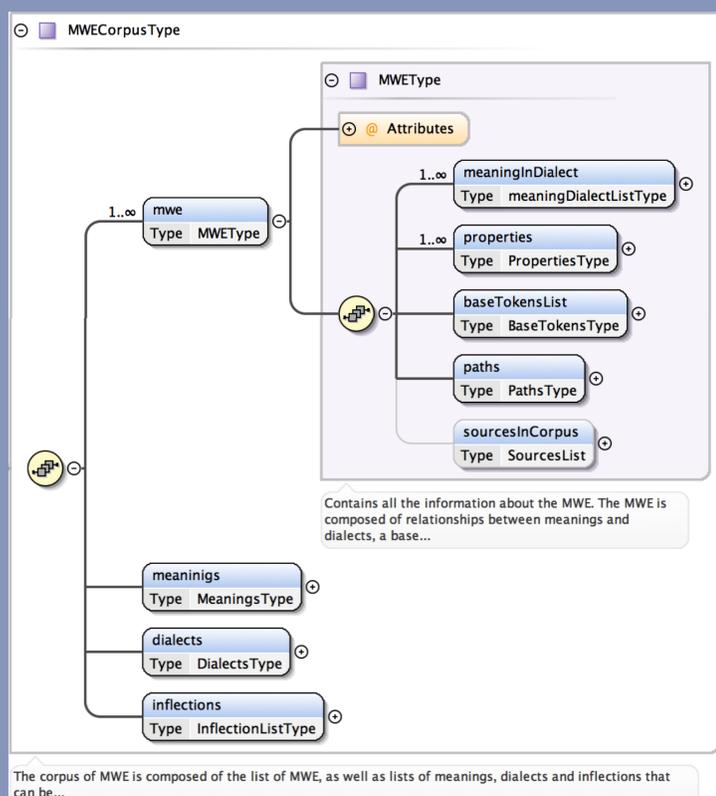
More Spanish dialects = more challenges

- ▶ Unique expressions
 - ▷ *Estar limpio_{CR}* [to be clean] → to be out of money
- ▶ Same meaning, different expression
 - ▷ *Estar aguja_{COL}* [to be needle] → to be out of money
- ▶ Same expression, same meaning
 - ▷ *Echar los perros_{COL,CR,MEX}* [to throw the dogs] → to flirt
- ▶ Same expression, different meaning
 - ▷ *Ponerse las pilas* [to put on the batteries]
 - To get one's act together_{COL}
 - To do things in a better way_{CR}
 - To be more active_{MEX}
 - To do things faster_{PE}



XML schema to represent Spanish MWEs

MWE definition within XML schema:



based on [1]

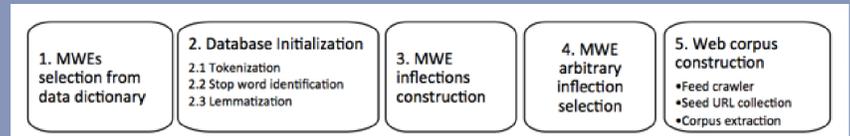
XML schema example

▶ *Hacerse el loco*:

```
<mwe id="MWE2" mweText="hacerse el loco" length="3">
  <meaningInDialect id="MWE2WPCO" meaning="WP" dialect="CO"/>
  <meaningInDialect id="MWE2WPCR" meaning="WP" dialect="CR"/>
  <meaningInDialect id="MWE2WPME" meaning="WP" dialect="MEX"/>
  <meaningInDialect id="MWE2WPPE" meaning="WP" dialect="PE"/>
  <properties>
    <allowsAdditions value="true"/>
    <allowsSubstitutions value="false"/>
    <allowsInflections value="true"/>
    <languageRegister value="Colloquial"/>
    <additionalTokensList [5 lines]>
  </properties>
  <baseTokensList id="MWE2_BTkn1">
    <baseToken id="MWE2_BTkn1" isStopWord="false" position="1" allowsInflections="true"
      allowsSubstitutions="false" analysis="V.W">
      <wordES>hacer</wordES>
      <wordEN>become</wordEN>
    </baseToken>
    <baseToken id="MWE2_BTkn2" isStopWord="true" position="2" allowsInflections="true"
      allowsSubstitutions="false" analysis="PRO.3s">
      <wordES>se</wordES>
      <wordEN>itself</wordEN>
    </baseToken>
    <baseToken id="MWE2_BTkn3" isStopWord="false" position="2" allowsInflections="true"
      allowsSubstitutions="false" analysis="DET.ms">
      <wordES>el</wordES>
      <wordEN>the</wordEN>
    </baseToken>
    <baseToken id="MWE2_BTkn4" isStopWord="false" position="3" allowsInflections="true"
      allowsSubstitutions="false" analysis="N.ms">
      <wordES>loco</wordES>
      <wordEN>crazy</wordEN>
    </baseToken>
  </baseTokensList>
  <paths>
    <path dialects="CO CR MEX PE"> [17 lines]
    <path dialects="CO CR MEX PE"> [18 lines]
    <path dialects="CO CR MEX PE"> [19 lines]
  </paths>
</mwe>
```

Lexical Resource and Corpus

- ▶ 250 MWE examples were gathered, natives speakers of the four Spanish dialects have described 40 of them.
- ▶ Methodology created to construct a Corpus based on web, founded in 5 steps:



MWEs examples from corpus

Extraction of 3 types of MWE examples from the corpus, aimed by a crowdsourcing strategy:

- ▶ **Positive example:** a given MWE occurs with its idiomatic meaning
 - ▷ *Pienso gastarme hasta el último peso, espero antes de colgar los zapatos.*_{MEX} [Think spend-me until the last penny, hope before of hang the shoes] → I plan to spend until my last penny, hopefully before I die
- ▶ **Neutral example:** a given MWE has a literal (compositional) meaning.
 - ▷ *Una idea interesante es modificar una percha de alambre para colgar los zapatos y, de esta manera, ahorrar espacio.* [An interesting idea is to modify a metal hanger to hang shoes and, in this way, save space] → An interesting idea is to modify a metal hanger to hang shoes; on this way you will save space.
- ▶ **Negative example:** all lexicalized components of the MWE occur but their syntactic dependencies are not those assumed by the MWE corpus examples.
 - ▷ *Los famosísimos zapatos de la serie Sexo en Nueva York ahora están disponibles para comprarlos y colgártelos.*_{MEX} [The famous shoes of series Sex in New York now are available for purchase and wear] → The very famous shoes of the Sex and the City TV show are now available to buy and wear.

MWEs support image extraction

- ▶ Creation of the basis of an image database to relate an illustration to the MWE example, to support the MWE understanding.
 - ▷ *Colgar los zapatos*_{MEX} [hang the shoes] → to die



Bibliography

- [1] Alon Itai and Shuly Wintner. Language resources for hebrew. *Language Resources and Evaluation*, 42(1):75–98, 2008.