

Uso de PLN en otras disciplinas

Gemma Boleda

GPLN

Universitat Politècnica de Catalunya

Jornadas PLN-TIMM. Colmenarejo, 4-6 febrero 2009.

Uso de PLN en otras disciplinas

Human knowledge is expressed in language. *So computational linguistics is very important.*

Mark Steedman, On becoming a discipline
Comp. Ling., 34(1), 2008.

- muchas disciplinas usan textos digitalizados para explorar
 - propiedades del lenguaje expresado en textos (Física, Lingüística)
 - conocimiento expresado en textos (Ciencias Sociales)
- **no usan PLN!**
- algunas excepciones:
 - BioNLP, Opinion Mining, Information Retrieval, lingüística de corpus . . .
 - (pero “internas” al PLN/LC)

Uso de PLN en otras disciplinas

Human knowledge is expressed in language. *So computational linguistics is very important.*

Mark Steedman, On becoming a discipline
Comp. Ling., 34(1), 2008.

- muchas disciplinas usan textos digitalizados para explorar
 - propiedades del lenguaje expresado en textos (Física, Lingüística)
 - conocimiento expresado en textos (Ciencias Sociales)
- **no usan PLN!**
- algunas excepciones:
 - BioNLP, Opinion Mining, Information Retrieval, lingüística de corpus ...
 - (pero “internas” al PLN/LC)

Uso de PLN en otras disciplinas

Human knowledge is expressed in language. *So computational linguistics is very important.*

Mark Steedman, On becoming a discipline
Comp. Ling., 34(1), 2008.

- muchas disciplinas usan textos digitalizados para explorar
 - propiedades del lenguaje expresado en textos (Física, Lingüística)
 - conocimiento expresado en textos (Ciencias Sociales)
- **no usan PLN!**
- algunas excepciones:
 - BioNLP, Opinion Mining, Information Retrieval, lingüística de corpus ...
 - (pero "internas" al PLN/LC)

Uso de PLN en otras disciplinas

Human knowledge is expressed in language. *So computational linguistics is very important.*

Mark Steedman, On becoming a discipline
Comp. Ling., 34(1), 2008.

- muchas disciplinas usan textos digitalizados para explorar
 - propiedades del lenguaje expresado en textos (Física, Lingüística)
 - conocimiento expresado en textos (Ciencias Sociales)
- **no usan PLN!**
- algunas excepciones:
 - BioNLP, Opinion Mining, Information Retrieval, lingüística de corpus . . .
 - (pero “internas” al PLN/LC)

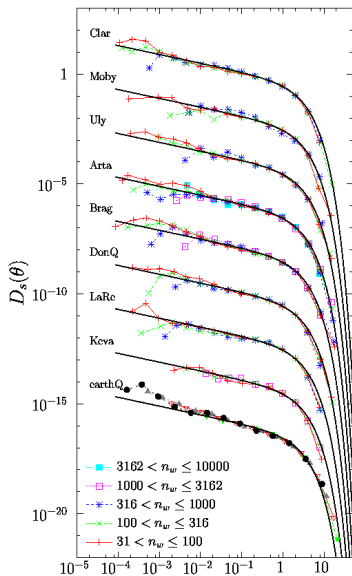
Ejemplo: distribución de distancias entre palabras

- Trabajo en colaboración con Álvaro Corral, Ramon Ferrer i Cancho, Albert Díaz-Guilera.
- Física: estudio de la distribución de distancias entre las ocurrencias de una misma palabra
 - El gato come **pescado**. A mí no me gusta el **pescado**.
 - distancia $l = 7$. Normalización: $\theta = \frac{1}{l}$
 - 10 novelas en 4 lenguas
- trabajos relacionados del mismo campo no utilizan taggers ni lematización

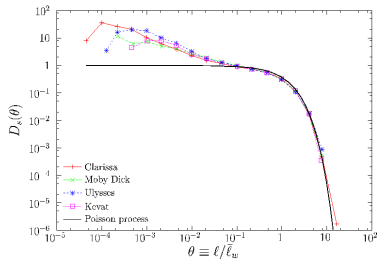
Ejemplo: distribución de distancias entre palabras

- Resultados preliminares:
 - la distribución de las palabras no es uniforme: todas presentan *burstiness* [Church y Gale 1995], incluso las funcionales (determinantes, preposiciones)
 - las palabras retóricamente importantes presentan más *burstiness* que el resto.
 - separación en vocabulario básico y específico mediante algoritmo de detección de comunidades

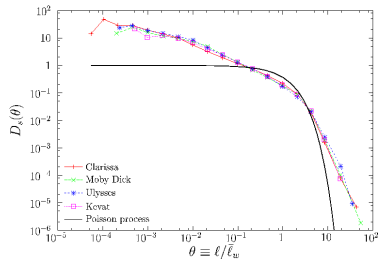
Ejemplo: distribución de distancias entre palabras



Ejemplo: distribución de distancias entre palabras



(a) Common vocabulary



(b) "Special" vocabulary

Conclusiones

- Interdisciplinariedad!
 - Física: leyes universales, estudio de sistemas complejos, redes
 - Psicolingüística: memoria, *priming*
 - Lingüística: clases de palabras
 - Retórica, Análisis Literario, Análisis del Discurso
 - ... y PLN
- Dificultades
 - robustez, precisión, adaptabilidad
 - 'The' nombre propio
 - 'Stephen' adverbio
 - difusión, pedagogía