Alumna: Guillermina Pagani

Asignatura: Análisis computacional de datos lingüísticos

Profesor: Javier Vera Zúñiga

**Proyecto de investigación**

¿Qué datos pretende utilizar?

Corpus del español disponibles online:

* CORLEC (Corpus Oral de Referencia de la Lengua Española Contemporánea) <http://www.lllf.uam.es/ING/Corlec.html>
* PRESEEA (Corpus del Proyecto para el estudio sociolingüístico del español de España y de América)  <http://preseea.linguas.net/Corpus.aspx>
* El Corpus del Español [www.corpusdelespanol.org](http://www.corpusdelespanol.org/)
* Codicach (Corpus Dinámico del Castellano de Chile) <http://sadowsky.cl/codicach-es.html>. Si fuera posible, también quisiera trabajar con el Corpus Oral Sociolingüístico del Castellano de Chile (COSCACH)

Adicionalmente, quisiera incorporar datos de redes sociales.

¿Qué técnicas computacionales espera utilizar para analizar los datos?

1. Uso de expresiones regulares/variables en Python y de identificación de palabras siguientes/anteriores.
2. Uso de tagging, tal vez, para evaluar las combinaciones posibles con *cagarse*. Uso de los rasgos de [+/− dinamicidad], [+/− delimitación] y [+/− duratividad], y de alguna forma de identificar los sujetos.

¿Qué problemas lingüísticos pretende abordar?

La construcción *cagarse* + gerundio, del registro coloquial del español de Argentina, está formada por el verbo flexionado *cagar*, un pronombre personal concordante en persona y número con este, y un gerundio; y parece codificar diferentes valores semánticos:

1. *Me cagué trabajando* *para comprar mi terreno.* (=trabajé mucho)
2. *La vegana del gym se cagó desmayando.* (=se desmayó bien desmayada)

Lo que quisiera investigar es lo siguiente:

1. En qué países se documenta la construcción. Para Chile parece haber una similar, que se podría agregar al relevamiento para realizar un contraste: *hacerse mierda* + gerundio, con valor cuantificativo.
2. Cuál es su funcionamiento sintáctico y semántico. En particular, mediante el análisis computacional, quisiera investigar: a) si su uso está aspectualmente restringido. Para ello, necesito determinar el aspecto léxico (*aktionsart*) de los verbos con los que se utiliza. Como señala la bibliografía, parece existir una relación entre la posibilidad de cuantificación de un predicado y la presencia del rasgo [+/− delimitado]. b) El tipo de sujetos con los que se combina (agentes, pacientes/temas, etc.).