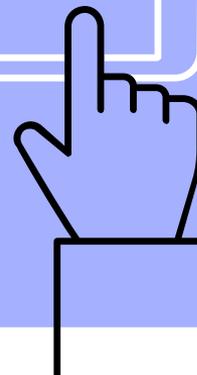
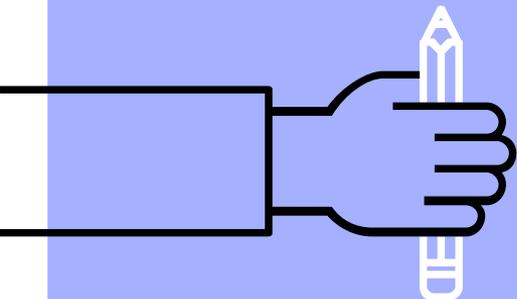
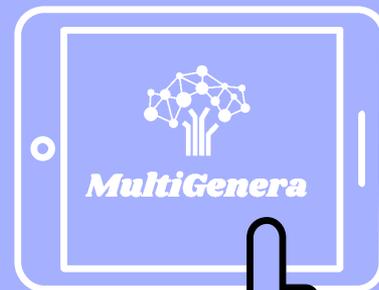
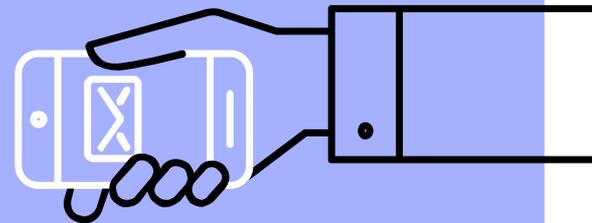
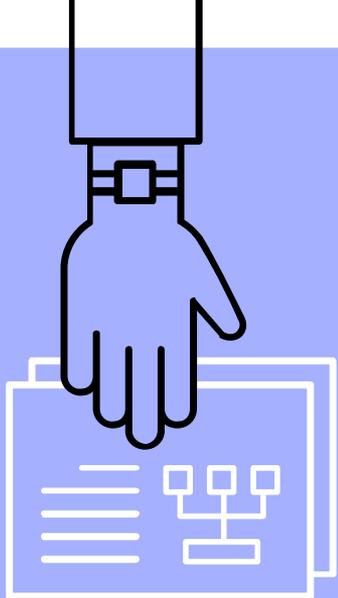


PROYECTO MULTIGENERA

Generación automática de ejemplos a
partir del diccionario de valencias
multilingüe PORTLEX

M^a José Domínguez Vázquez
Carlos Valcárcel Riveiro

Fundación **BBVA**



¡Hola!

Somos Majo y Carlos

Trabajamos juntos en
muchos proyectos.

@CValcrcel

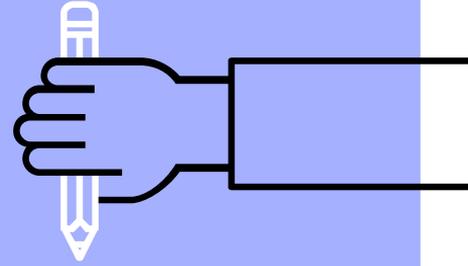
@majodominguez13



CONTENIDOS

1. ¿Qué hacemos y para qué?
2. El equipo.
3. El proyecto.
4. Herramientas.
5. Generar un complemento.

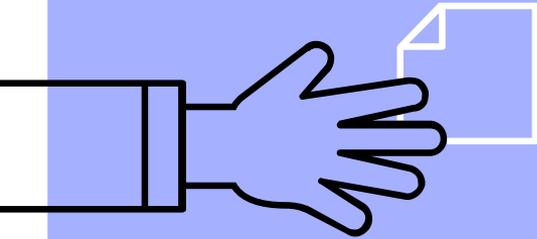


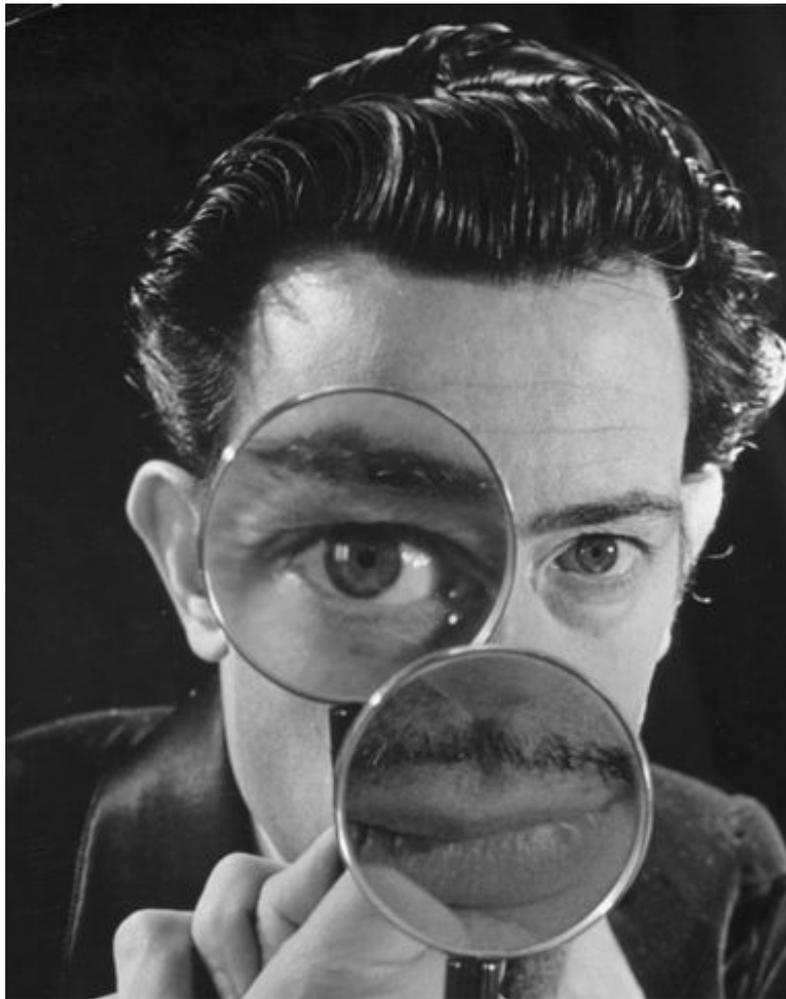


1.

MULTIGENERA

¿Qué hacemos y para qué?





Pasamos
mucho tiempo
buscando
ejemplos

NECESITAMOS GENERAR EJEMPLOS DIDACTIZADOS PARA...



Diccionarios

Diccionarios escolares.

Diccionarios de aprendizaje.

Diccionarios valenciales.



Enseñar

Explicaciones gramaticales.

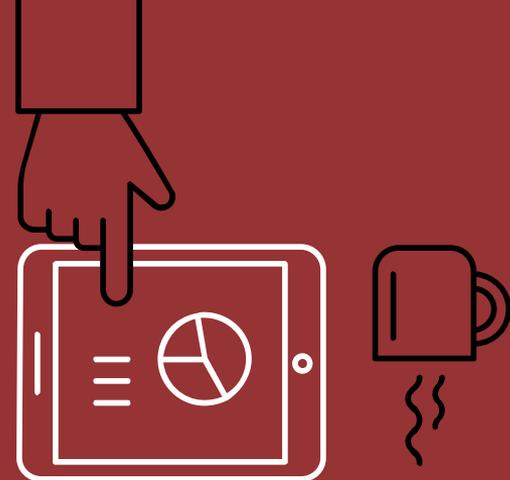
Explicaciones de léxico y semántica.



Aprender

Ejercicios estructurales (focus on language).

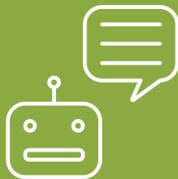
Actividades de deducción gramatical.





OBJETIVOS

Diseñar un prototipo para generar automáticamente frases nominales en tres lenguas.



Generar frases nominales

Estructuras argumentales de sustantivos valenciales.

Mono, bi y triactanciales

3 lenguas: castellano, francés y alemán.



Desarrollar un prototipo

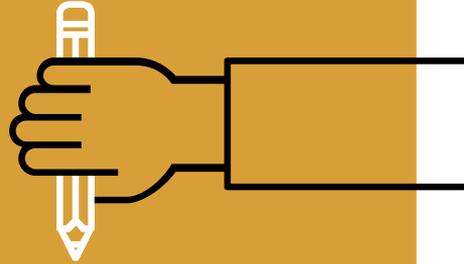
10 lexemas.

Optimización de recursos preexistentes: diccionarios, corpus, nets.

Acceso abierto.

El olor a tabaco

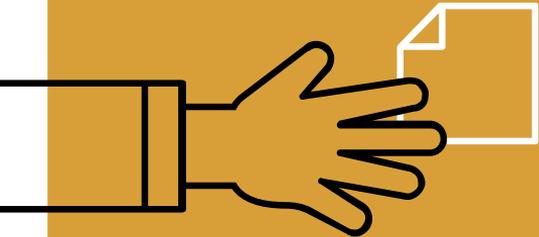




2.

EL EQUIPO

¿Quiénes somos...
y de dónde venimos?



Un equipo multidisciplinario



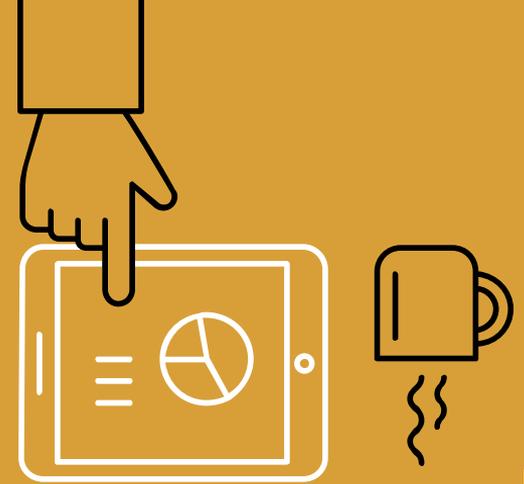
Diferentes **países** y universidades



Diferentes **lenguas** de especialidad



Diferentes **disciplinas**



Un equipo, dos grupos de trabajo



Mª José Domínguez

(USC)

IP del proyecto

Coordinación general



Mª Teresa Sanmarco

(USC)

Coordinación del
grupo de trabajo
lingüístico.



Carlos Valcárcel

(UVigo)

Coordinación del
grupo de trabajo
computacional.



GRUPO DE COMPUTACIÓN



Isaac González López

(Cilenis / Xunta)

Programación.



David Lindemann

(Universität
Hildesheim)

Corpus, Wordnet.



Carlos Valcárcel

(UVigo)

Coordinación de
grupo.



Miguel A. Solla

(RELEX / UVigo)

Asesor externo:
Corpus, Wordnet.



Paulo Gamalho

(Cilenis / USC)

Asesor externo:
Corpus, Wordnet.

GRUPO DE LINGÜÍSTICA



Ma Teresa Sanmarco

(USC)

Coordinación, web y
PORTLEX



Laura Pino Serrano

(USC)

Francés



Carlos Valcárcel

(Uvigo)

Francés y PORTLEX



Sascha Wolfer

(IDS - Mannheim)

Gestión de datos.



Rosa Martín

(UCM)

Castellano y web



Alberto Bustos

(UEX)

Castellano y francés



Natàlia Català

(U. Rovira i Virgili)

Castellano



José A. Calañas

(U. de València)

Alemán



Mónica Mirazo

(USC)

Alemán y PORTLEX



Carolin Müller-Spitzer

(IDS - Mannheim)

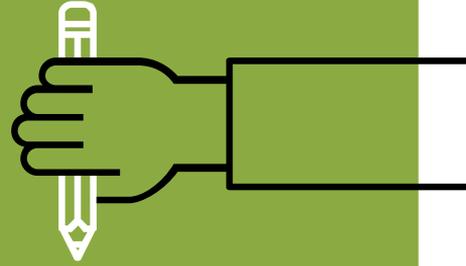
Alemán



María José Domínguez

(USC)

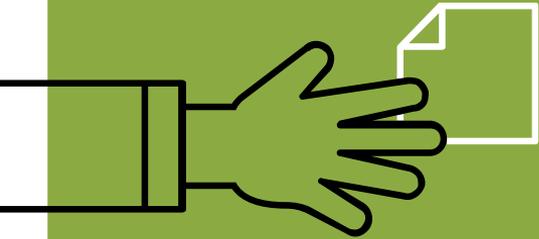
Alemán y PORTLEX



3.

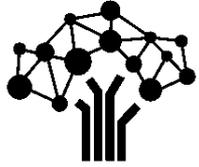
EL PROYECTO

Fundamentos y gestión.



3 PROYECTOS ENCADENADOS





MultiGenera

Fundación BBVA

Ayudas para la investigación en el
área de sociedad digital y
humanidades digitales

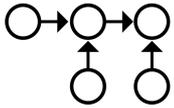
67299,19€

para dos años

María José Domínguez

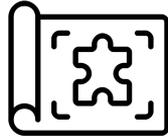
Investigadora principal. USC.

FUNDAMENTOS TEÓRICOS



Gramática valencial

Valencia
Actante / argumento
Rol semántico
Rasgo categorial
Estructura argumental / Combinatoria

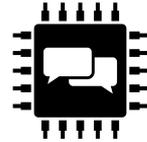


Teoría de prototipos

Prototipos léxicos.

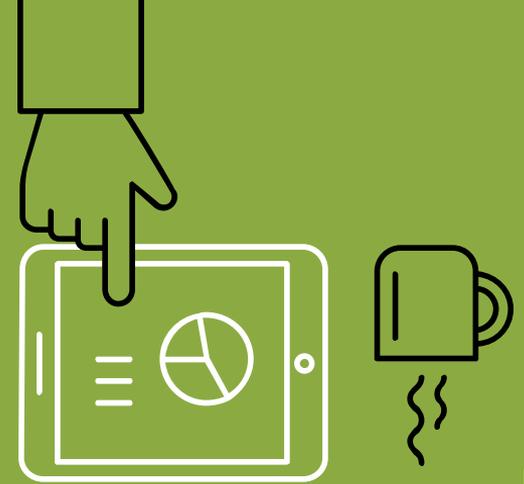


Clases semánticas.



Procesamiento del lenguaje natural

Recuperación de información.
Extracción de información.
Generación del lenguaje natural.



OLOR

"Emanación que producen los cuerpos y que se percibe por el sentido del olfato"
CLAVE

ACTANTE 1

Aquel o aquello que tiene o dispone de algo

[+Material]
[+Animado]
[+Situación, +Lugar]
[+Situación, +Tiempo]

ACTANTE 2

Clasificativo

[+Material, +Sustancia]
[+Material, +Objeto]
[+Animado, +Planta]
[+Animado, +Animal]
[+Situación, +Estado]



De + det + sustantivo

Adjetivo

A + sustantivo

De + sustantivo

Adjetivo

ROL

RASGOS

REALIZACIONES



El olor de la habitación

El olor corporal

El olor a limón

El olor de naranjas

El olor fecal

El olor a café de la cocina

El olor de tus manos a incienso

El olor de gasolina del garaje

El olor resinoso de aquella habitación



 No se encuentran en el corpus

?El olor corporal a sudor

?El olor masculino de loción

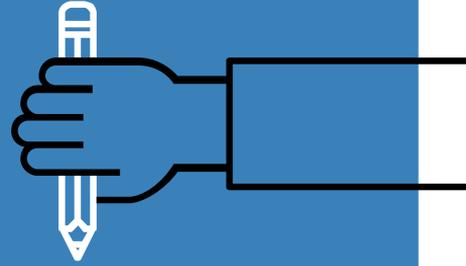
?El olor corporal fecal



*El olor a cristal

*El olor de la madera resinoso

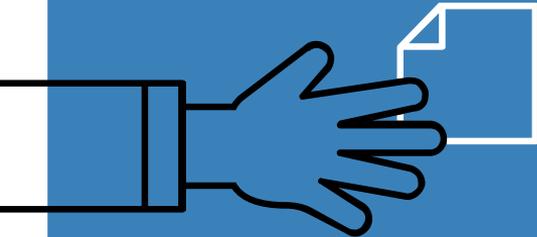
*El olor del consejo a azúcar



3.

HERRAMIENTAS

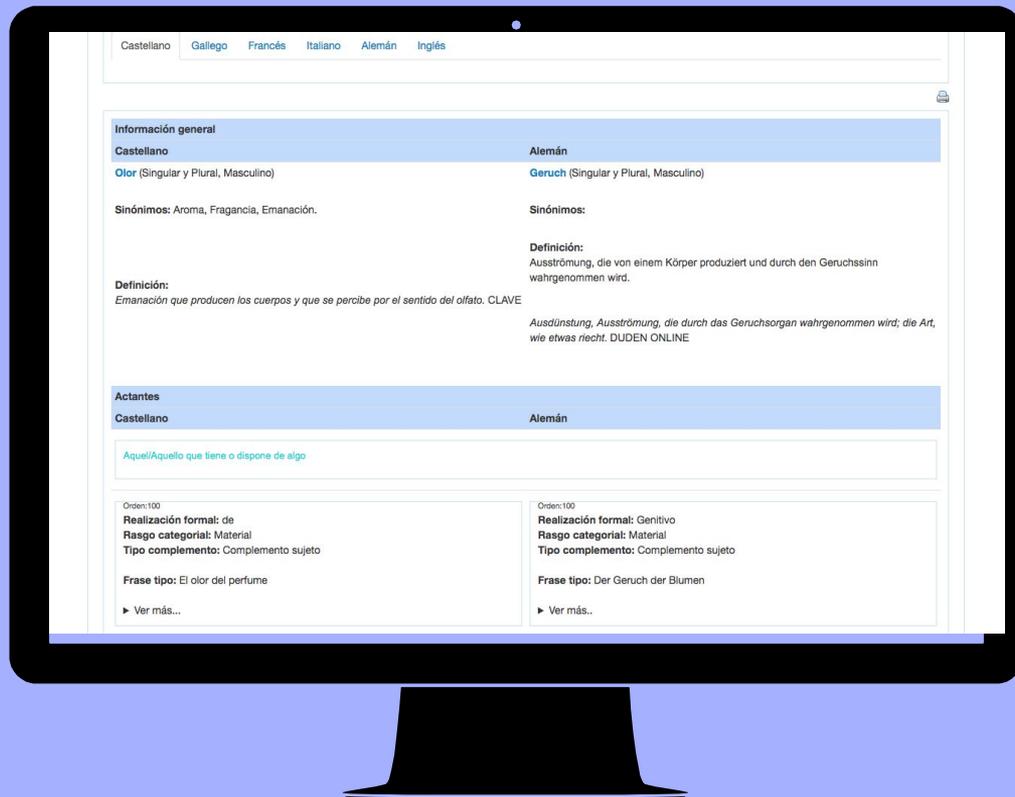
Integración y optimización.



PORTLEX

Estructuras
argumentales.

Rasgos
semánticos.



CORPUS

CREA, DeReKo,
FRANTEXT...

Pero sobre todo
SKETCH ENGINE.

Extracción de
prototipos léxicos.

Sketch Engine Spanish Web 2011 (esTenTen11, Eu + Am) freq = 213.667 (19.44 per million)

olor (noun)
Spanish Web 2011 (esTenTen11, Eu + Am) freq = 213.667 (19.44 per million)
displaying only: prepositional phrases whole word sketch

"olor" a ...	21.62	"olor" de ...	13.24	... de "olor"	11.65	... con "olor"	5.75	... por "olor"	2.46
azufre + 298 9.09 olor a azufre		napalm + 379 8.72 Me encanta el olor del Napalm por la mañana		clavo + 2,942 11.59 clavos de olor		impregnar + 143 8.15 mierda + 105 7.88 la misma mierda con distinto olor .		molestia 92 8.22 las molestias por olores	
quemado + 795 9.09 olor a quemado		santidad + 318 7.88 en olor de santidad		hierba + 400 8.59 las hierbas de olor		humo 87 7.71 de humo con olor		atraer + 281 7.97 quejar + 107 7.53 vecino 70 7.53	
pólvora + 770 9.05 el olor a pólvora		perfume + 287 7.76 olor del perfume		impregnar + 541 8.59 emanación + 351 8.49 emanación de malos olores		perfume 75 7.52 un perfume con olor a		impregnar 30 7.05 embriagar 21 6.79 contaminación 70 6.78 contaminación por olores	
gas + 1,169 8.99 olor a gas		incienso + 147 7.25 el olor del incienso		roble + 174 7.69 Roble de olor		aliento 69 7.45 aliento con olor a		alerter 41 6.68 nariz 19 6.66 perfumar 16 6.56 marear 16 6.42 mareado por el olor	
humedad + 785 8.93 olor a humedad		multitud + 218 7.10 en olor de multitudes		quejar + 219 6.45 neutralizador 65 6.39 neutralizador de olores		orina 65 7.34 de la orina con olor a		alerter 41 6.68 nariz 19 6.66 perfumar 16 6.56 marear 16 6.42	
pescado + 662 8.75 el olor a pescado		pólvora + 153 6.94 el olor de la pólvora		sustancia + 129 6.29 sustancias de olor		aire con olor a		invadir 55 6.26 queja 45 6.20 despertar 25 6.17 mareo 12 6.16 inundar 21 6.09 excitar 14 6.01 desmayar 11 5.93 detectar 49 5.92 alerta 29 5.68	
humo + 707 8.67 olor a humo		sudor + 134 6.79 olor del sudor		percatar + 127 6.25 líquido + 105 6.21 líquido de olor		flujo 69 7.29 flujo con olor		La alerta por el olor es insuficiente .	
tabaco + 604 8.42 el olor a tabaco		orina + 181 6.77 el olor de la orina		trébol 57 6.14 trébol de olor		ropa 71 7.03 la ropa con un olor		alarmar 14 5.65 molesto 20 5.61 asco 9 5.60 vomitar 8 5.50 delatar 8 5.43	
guayaba + 88 6.50 El olor de la guayaba		incienso + 478 8.38 el olor a incienso		flor + 340 6.38 el olor de las flores		perfumar 54 6.89 sustancia 54 6.67 embriagar 40 6.60 veta 43 6.57 velas con olor a vainilla ,			
alcohol + 500 8.13 fuerte olor a alcohol		alcohol + 500 8.13 fuerte olor a alcohol		heces 89 6.32 volumen y el olor de las heces		eliminar + 362 5.96 la eliminación de olores			
mierda + 527 8.13 el olor a mierda		incienso + 478 8.38 el olor a incienso		jazmin 78 6.30 olor de los jazmines		rastros + 108 5.88 rastros de olor			
carne + 486 8.11 olor a carne quemada		aliento + 122 6.21 olor del aliento		aliento + 122 6.21 olor de los jazmines		gusante 44 5.84 gusantes de olor			
naftalina + 393 8.10 con olor a naftalina		huevo + 417 7.98 olor a huevos podridos		huevo + 417 7.98 olor a huevos podridos		tinta + 132 6.12 bacteria 64 5.82			
sudor + 330 7.83 el olor a sudor		sudor + 330 7.83 el olor a sudor							

GALNET

Clasificación
semántica

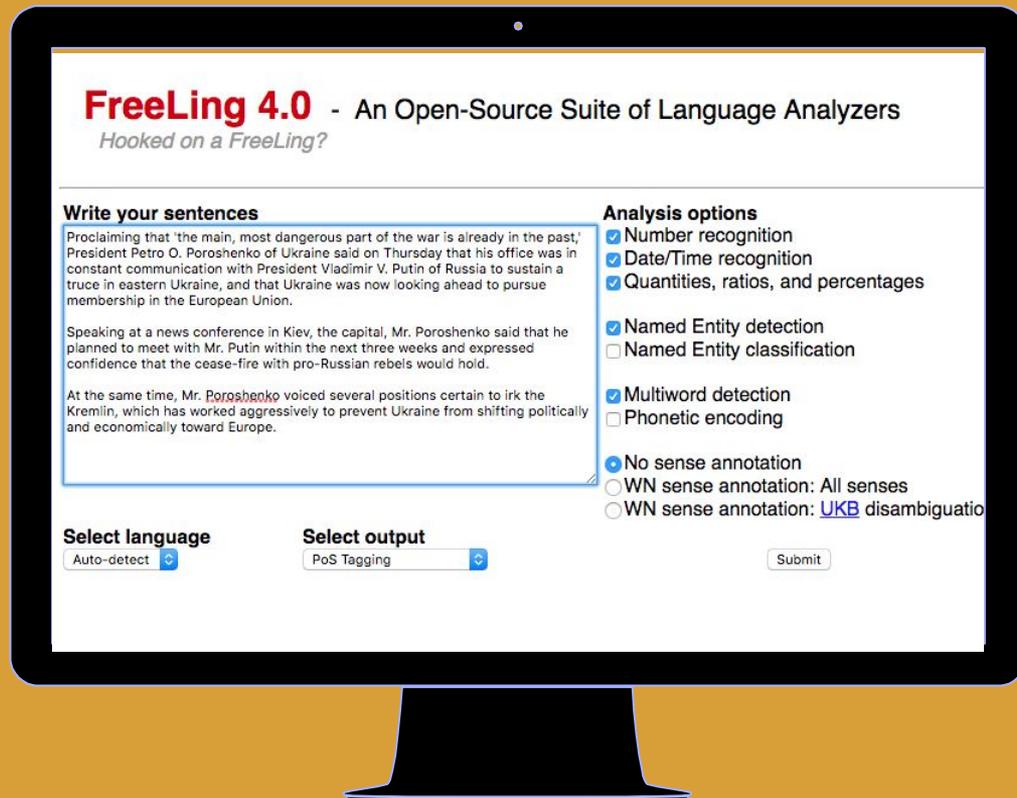
Ontologías

Expansión de
prototipos

+ café	n 1 [beverage]	spa-30-07929519-n	a beverage consisting of an infusion of ground coffee beans
+ café_con_hielo	n 1 [beverage]	spa-30-07920540-n	a strong sweetened coffee served over ice with cream
+ café_con_leche	n 1 [beverage]	spa-30-07919572-n	equal parts of coffee and hot milk
+ café_corto	n 1 [beverage]	spa-30-07919665-n	small cup of strong black coffee without milk or cream
+ café_exprés	n 1 [beverage]	spa-30-07919665-n	small cup of strong black coffee without milk or cream
+ café_helado	n 1 [beverage]	spa-30-07920540-n	a strong sweetened coffee served over ice with cream
+ café_instantáneo	n 1 [beverage]	spa-30-07920863-n	dehydrated coffee that can be made into a drink by adding hot water
+ café_irlandés	n 1 [beverage]	spa-30-07919441-n	sweetened coffee with Irish whiskey and whipped cream
+ café_moca	n 1 [beverage]	spa-30-07920872-n	a superior dark coffee made from beans from Arabia
+ cafe_noir	n 1 [beverage]	spa-30-07919665-n	small cup of strong black coffee without milk or cream
+ café_royal	n 1 [beverage]	spa-30-07929940-n	black coffee with Cognac and lemon peel and sugar
+ café_turco	n 1 [beverage]	spa-30-07921239-n	a drink made from pulverized coffee beans; usually sweetened
+ cafeto	n 1 [beverage]	spa-30-07929519-n	a beverage consisting of an infusion of ground coffee beans
+ cafe_latte	n 1 [beverage]	spa-30-07920222-n	strong espresso coffee with a topping of frothed steamed milk
+ cake	n 1 [cake]	spa-30-07028870-n	baked goods made from or based on a mixture of flour, sugar, eggs, and fat
+ calabacín	n 3 [veg]	spa-30-07716358-n	small cucumber-shaped vegetable marrow; typically dark green
+ calabacín_grande	n 1 [veg]	spa-30-07716504-n	large elongated squash with creamy to deep green skins
+ calabaza	n 1 [veg]	spa-30-07715561-n	edible fruit of a squash plant; eaten as a vegetable
+ calabaza	n 8 [veg]	spa-30-07716504-n	large elongated squash with creamy to deep green skins
+ calabaza	n 2 [veg]	spa-30-07735510-n	usually large pulpy deep-yellow round fruit of the squash family maturing in late summer or early autumn
+ calamar	n 2 [cookery]	spa-30-07781801-n	(Italian cuisine) squid prepared as food
+ calamares	n 1 [cookery]	spa-30-07781801-n	(Italian cuisine) squid prepared as food
+ calciferol	n 1 [aliment]	spa-30-15092227-n	a fat soluble vitamin that prevents rickets

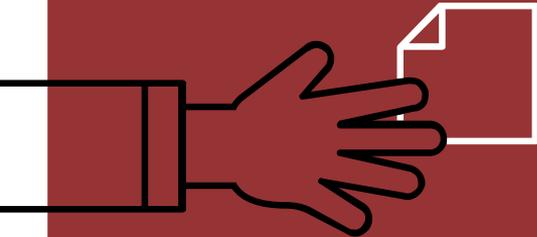
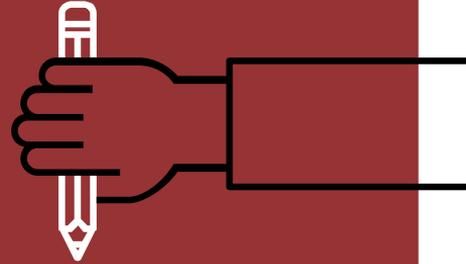
FREELING

Información
morfológica.
Formas
flexionadas.

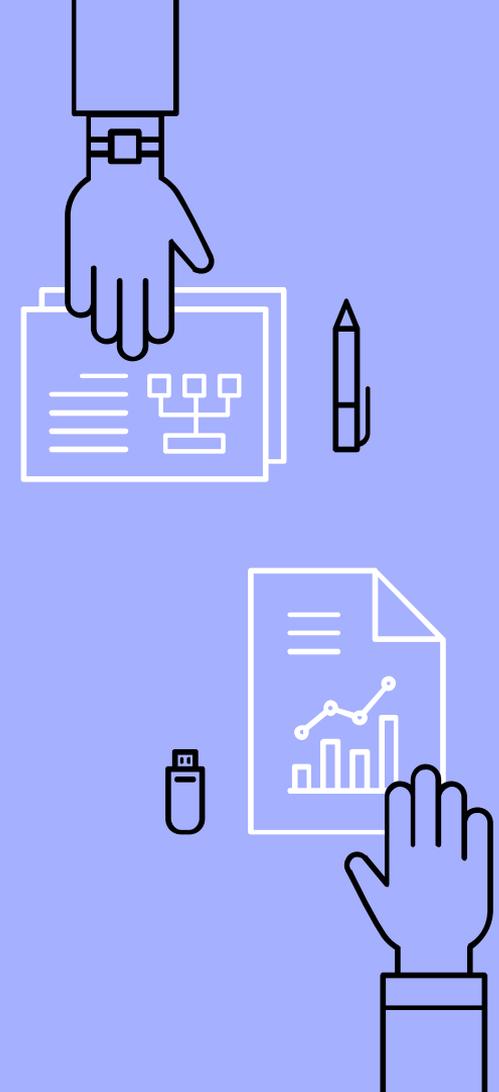
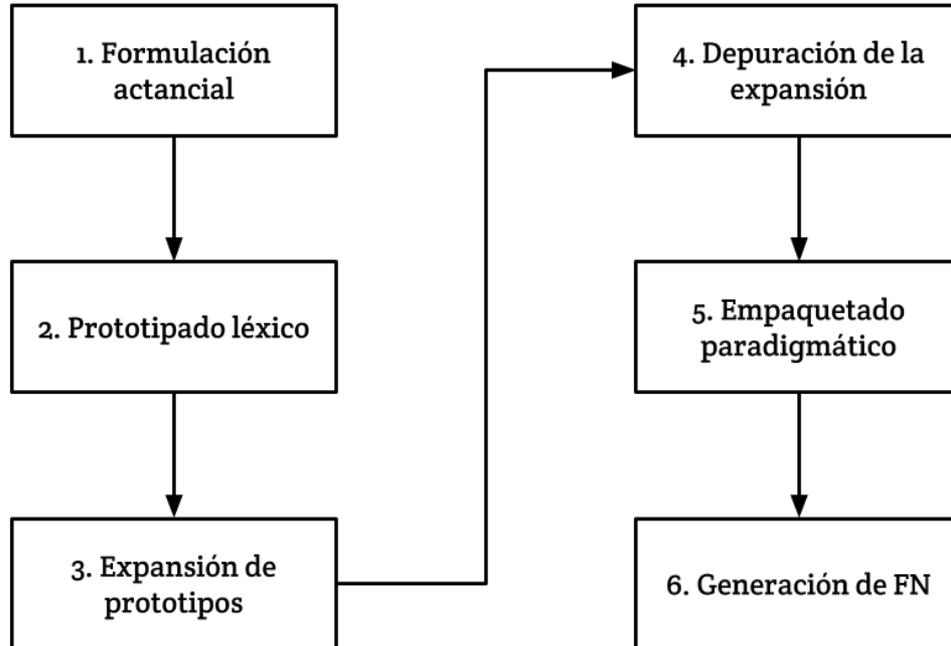


4. GENERAR UN COMPLEMENTO

¿Cómo lo hacemos?



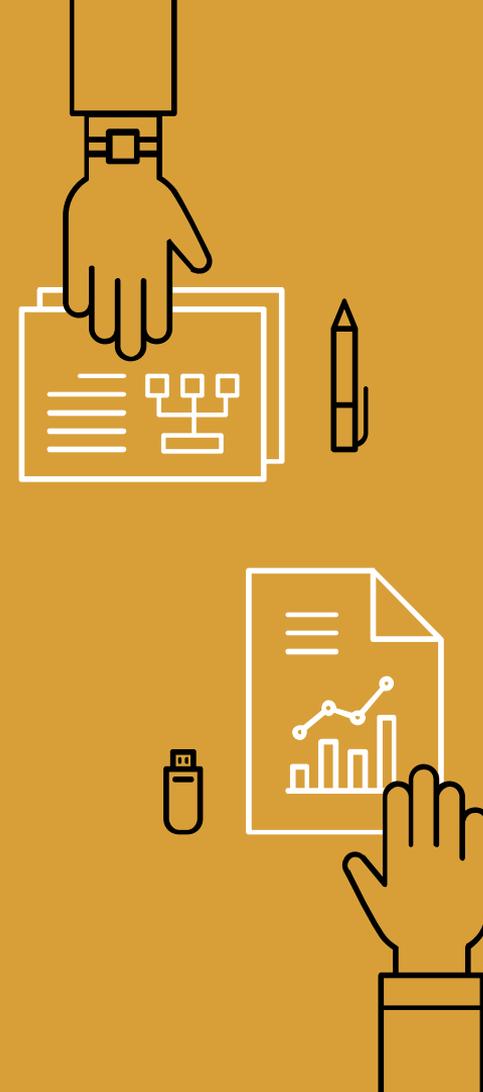
Fases de trabajo



Formulación actancial

Det.	{Adjetivo}	Núcleo de la FN	Det. Genitivo	Sust. A1 [Material]
Der El	angenehme <i>agradable</i>	Geruch <i>olor</i>	der <i>de las</i>	Blumen <i>flores</i>
Det.	{Adjetivo}	Núcleo de la FN	von (+ {Det.})	Sust. A1 [Material]
Der El	intensive <i>intenso</i>	Geruch <i>olor</i>	von diesen <i>de esos</i>	Männern <i>hombres</i>
Det.	{Adjetivo} A1	Núcleo de la FN	nach (+{Det.})	Sust. A2 [Material]
Der El	menschliche <i>humano</i>	Geruch <i>olor</i>	nach <i>a</i>	Schweiß <i>sudor</i>
Det.	{Adjetivo}	Adjetivo A1 [Animado]	Núcleo de la FN	
Der El	intensive <i>intenso</i>	männliche <i>masculino</i>	Geruch <i>olor</i>	
Det.	{Adjetivo}	Sust. A2 [Material]	Núcleo de la FN	
Der El	stechende <i>fuerte</i>	Schweiß <i>sudor</i>	-geruch <i>olor</i>	

Esquemas actanciales de GERUCH ('olor'). Fuente: PORTLEX.



Prototipado léxico. Pasos.

Det. + *olor a* + nombre común

Búsqueda en
corpus



Selección de
prototipos por
frecuencia



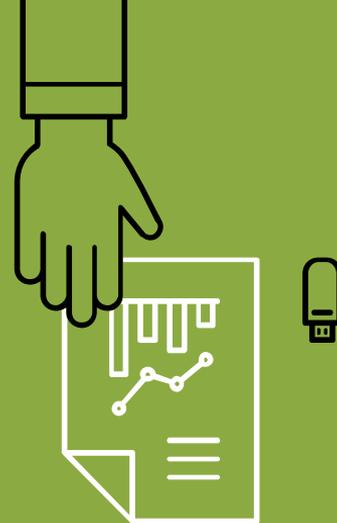
Clasificación
léxica



Prototipado léxico. A2 de OLOR

Prototipos léxicos	1º Orden	2º Orden	3º Orden	4º Orden
tabaco	Material	Sustancia	Sólido	Fumar
incienso	Material	Sustancia	Sólido	Químico
pólvora	Material	Sustancia	Sólido	Químico
humo	Material	Sustancia	Gas	Fumar
humedad	Situación	Estado	Propiedad	
gasolina	Material	Sustancia	Líquido	Carburante
azahar	Animado	Planta	Flor	
sudor	Material	Sustancia	Líquido	Excremento
azufre	Material	Sustancia	Element	Químico
naftalina	Material	Sustancia	Sólido	Químico

Realización: olor a + sustantivo. Fuente: Sketch Engine.



Tres clases semánticas centrales

[+Material]
[+Sustancia]

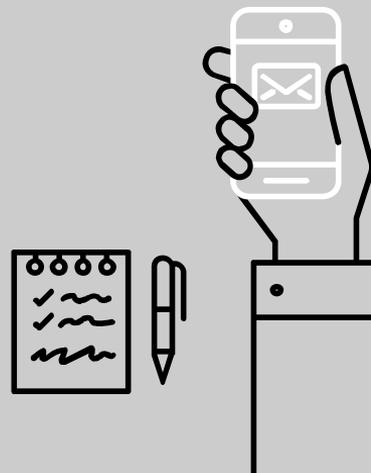
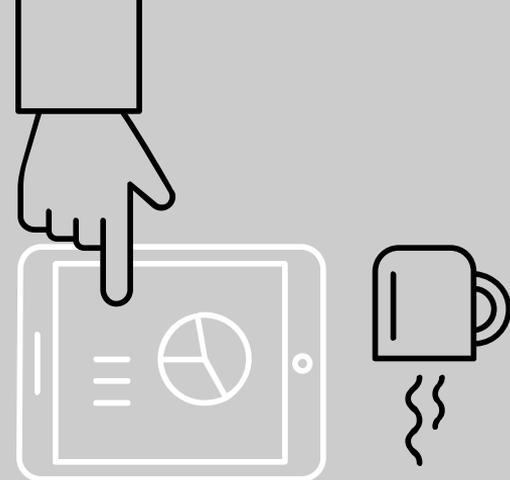
sudor
tabaco
pólvora

[+Animado]
[+Planta]

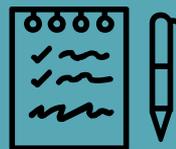
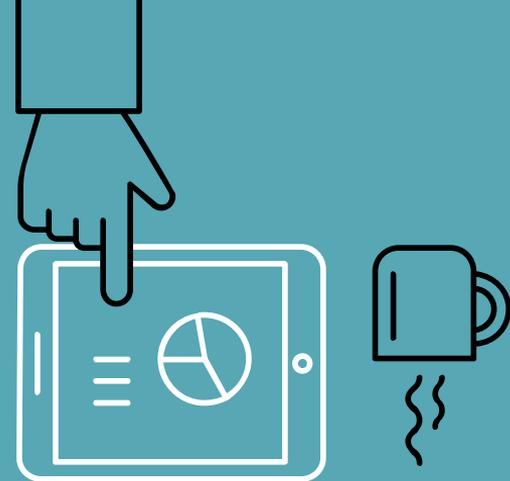
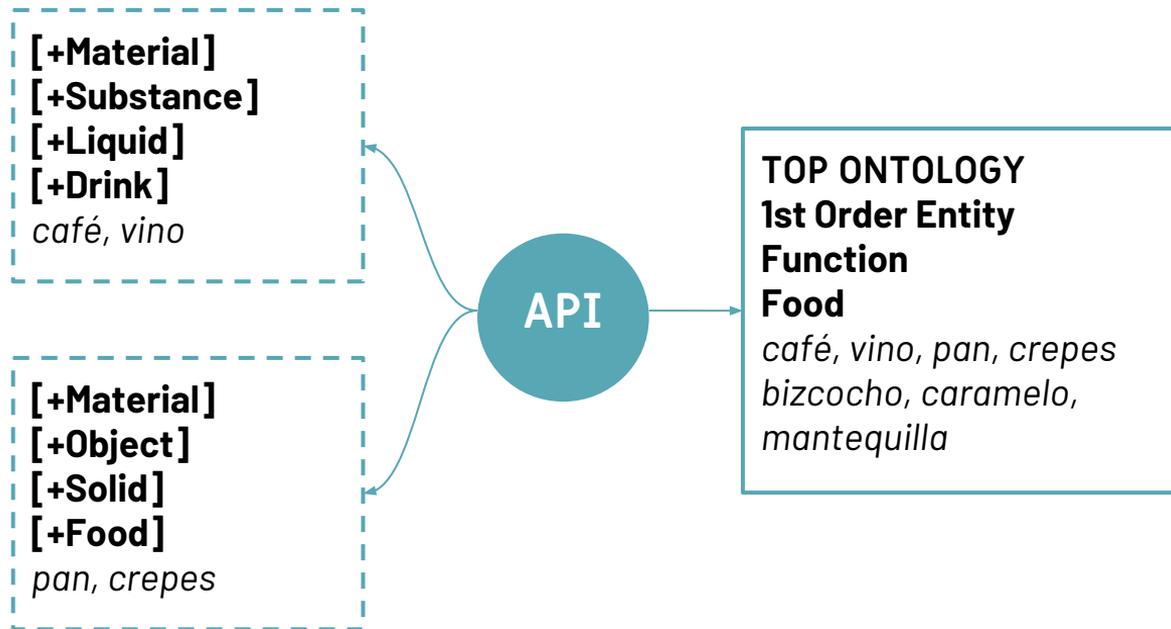
flor
rosa

[+Material]
[+Objeto]

pan
crepes



Extracción de lexemas para expandir listas de prototipos



Acercándonos a la generación automática

Depuración



Generación



Empaquetado



“



*It always seems
impossible until it's done*

Nelson Mandela

¡GRACIAS!

¿Preguntas?

Pueden escribirnos a:

carlos.valcarcel@uvigo.es

majo.dominguez@usc.es

Más información en:

multigenera.wordpress.com

#MultiGenera

R Σ LEX

