

Regular polysemy in context: a corpus-based analysis of regular polysemy in Spanish nouns

Irene Renau

Pontificia Universidad
Católica de Valparaíso

PHRASALEX
WORKSHOP, 22-23 July
2021

Fondecyt Project nr
1191204



**Agencia
Nacional de
Investigación
y Desarrollo**

Ministerio de Ciencia,
Tecnología, Conocimiento
e Innovación

Gobierno de Chile

01

Introduction
to the
project

02

Theoretical
background

03

Corpus-
based
methodology

04

Preliminary
results

05

Final remarks

Outline

Introduction to the project

Project Fondecyt nr 1191204 (ANID, Chile Government) 2019-2022 is devoted to:

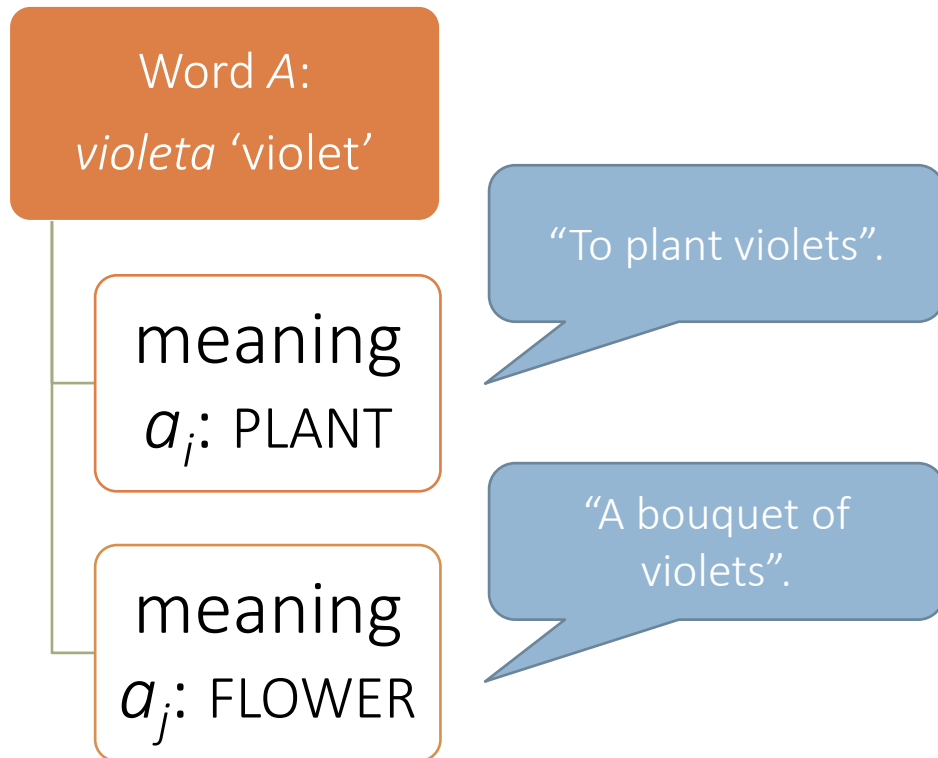
1. Conduct **corpus analysis** of highly frequent Spanish nouns to detect regular polysemy
2. **Classify** types of regular polysemy of Spanish nouns
3. Provide an open lexical **database** mapping corpus data with semantic description

Theoretical background

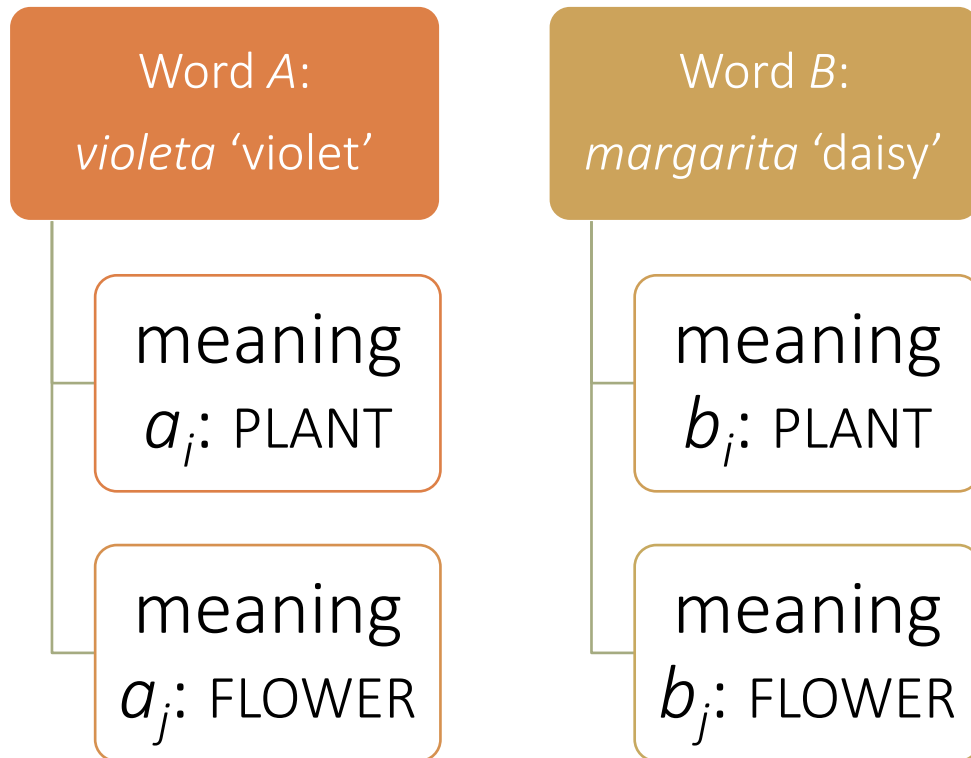
Regular polysemy is defined by Apresjan (1974) as follows:

“Polysemy of the word A with the meanings a_i and a_j is called regular if, in the given language, there exists at least one other word B with the meanings b_i and b_j , which are semantically distinguished from each other in the same way as a_i and a_j and a_i and b_j are nonsynonymous” (Apresjan 1974: 16).

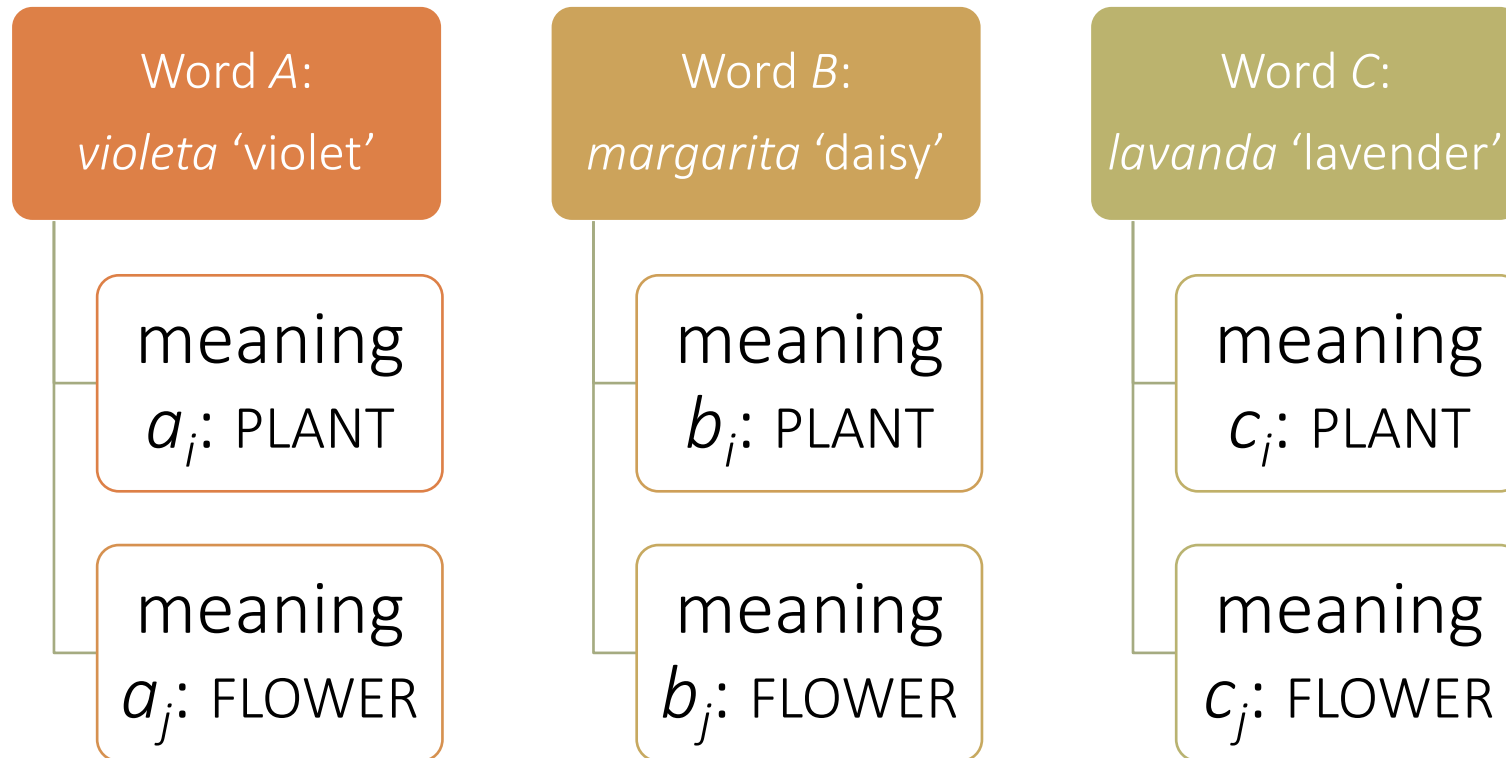
Theoretical background



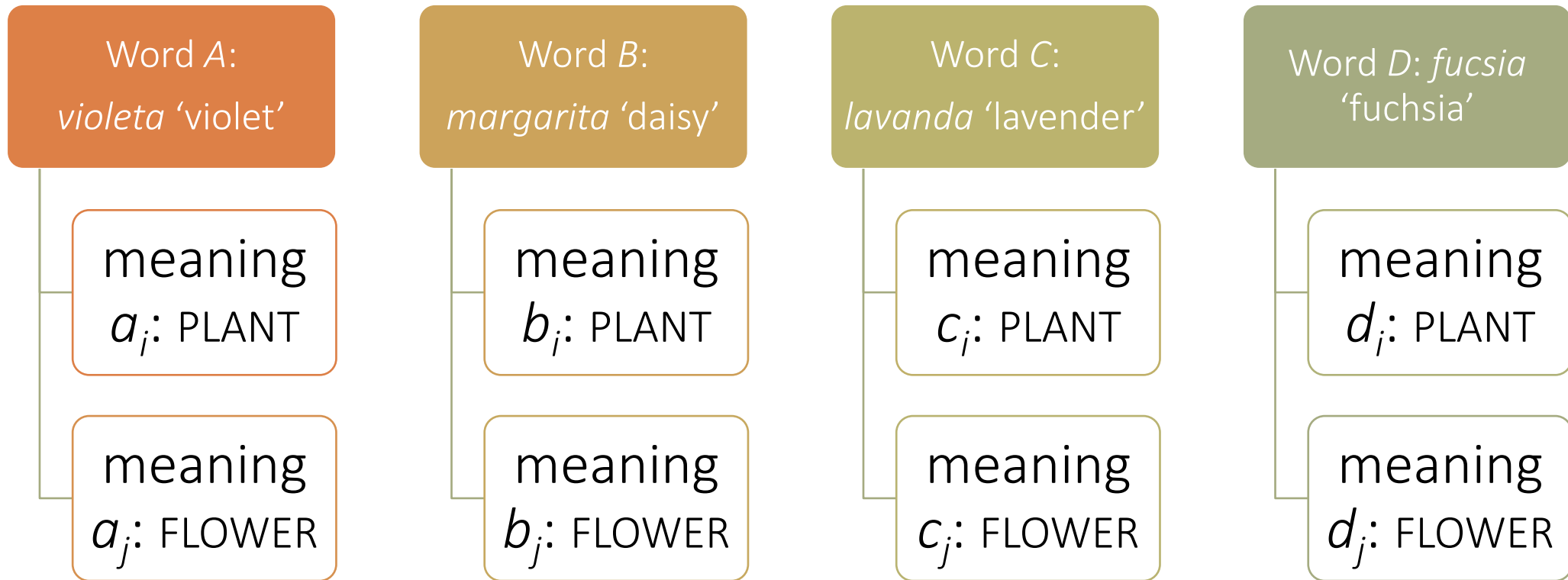
Theoretical background



Theoretical background



Theoretical background



Theoretical background

Regular polysemy is created by **metonymy**, e.g. “PLANT for PLANT PART”. Lakoff and Johnson (1980) explain that **conceptual metonymy** is a cognitive resource which can be found in the use of language.

Pustejovsky (1995) postulates that these pairs (e.g. PLANT / FLOWER) are packed together in the same semantic type. The context establishes if meaning a_i or a_j are used, or both (copredication, e.g. “a **heavy** and **boring book**”).

Differentiation between meanings is more **fuzzy and gradual** than in metaphor (e.g. “I bought a **book**” → PHYSICAL OBJECT / INFORMATION).

Theoretical background

Regular polysemy seems to be a **predictable** phenomenon due to its systematicity (Peters and Kilgarriff, 2000). E.g., one can predict that the document and the content will be called the same way (*book, novel, volume, film, trilogy, draft, etc.*).

The grade of **lexicalization and conventionalization** of the meaning is not predictable, as is usual in language (Rojas, 2011). E.g., *board* as 'content' (*read the board*)??

There are **extra-linguistic** constraints (cultural, technological...) for the stabilization of these pairs of meanings in the lexicon. E.g., a flower which is not very remarkable can have no name.

Theoretical background

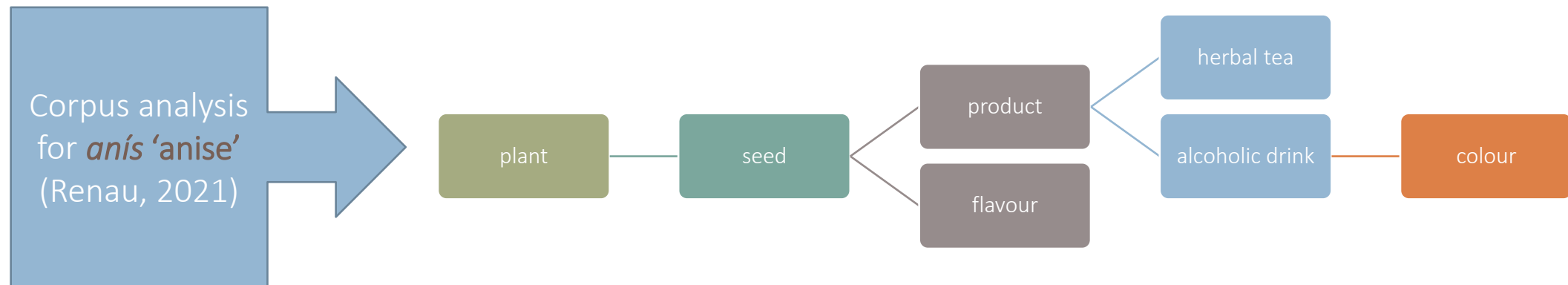
Dictionaries show the **variety** of representation of regular polysemy in different languages and between dictionaries of the same language (Renau, 2021):

		Spanish					Catalan	French	Italian	English	
Noun	Meaning	DEM	DIEA	DLE	DUE	Total Spanish	DIEC	TLF	Zing.	Coll.	Total dictionaries
<i>anís</i> 'anise'	alcoholic drink	1	1	1	1	4	1	1	0	0	6
	colour	0	0	0	0	0	0	0	0	0	0
	fruit	1	0	0	0	1	1	1	1	0	5
	herbal tea	0	0	0	0	0	0	0	0	0	0
	plant	1	1	1	1	4	1	1	1	1	8
	product	0	0	1	1	2	1	1	0	0	6
	seed	0	1	1	0	2	0	1	1	0	4
<i>frutilla</i> 'strawberry'	colour	0	0	0	0	0	0	0	0	0	0
	flower	0	0	0	0	0	0	1	0	0	1
	fruit	0	1	1	1	3	1	1	1	1	7
	plant	0	1	0	0	1	0	0	1	0	2

Theoretical background

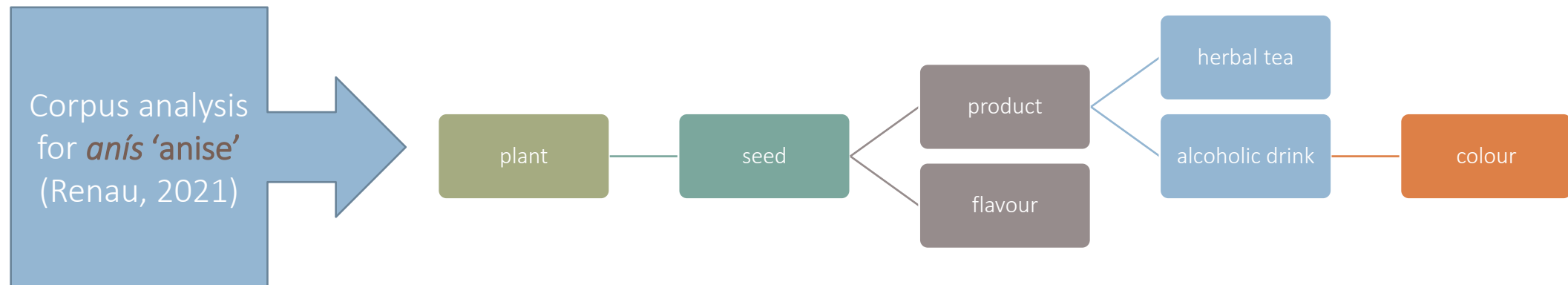
Most of the data for the study of regular polysemy in previous literature come from **introspection** or **secondary sources** (mainly Wordnet). Lack of direct sources of data is a problem to define the phenomenon with an empirical basis.

Some exceptions (**corpus-based approaches**): Goossens (2012), Ježek and Vieu (2014), Berri (2014), Ramírez (2020), Renau (2021). Corpus-based approaches show the complexity and richness of this phenomenon:



Theoretical background

Not very obvious how to structure this complexity in the **microstructure** of the pedagogical dictionary, preserving the needs of the learner (easy to understand, avoiding verbosity, etc.)!



Annotation mode is active

ANNOTATION ACTIONS

MANAGE ANNOTATIONS

Left context	KWIC	Label	Right context
El sexto año se ha producido una fuerte caída del precio del	café	semilla	y del azúcar, de modo que el Estado brasileño disminuye m
añana, y casi cada tarde: las medias lunas acompañadas de	café	infusión	con leche. </s><s> Nuestra guía particular (Hanna) nos hizi
versar por una semana aceptó salir conmigo y fuimos por un	café	infusión	. </s><s> Fue la cita perfecta, hablamos mucho, reímos y h
tarea y a finales de ese año nos llamaron a decirnos que el	café	semilla	había ganado un premio en un concurso internacional de ca
l Instituto Dominicano del Café. </s><s> Los productores de	café	semilla	denunciaron que tras de este proyecto de ley se oculta el p
as, sesiones paralelas, puestos de información y recesos de	café	infusión	y almuerzo para interactuar y ponerse al tanto de las noved
aración de las dos superficies abrasivas, entre los cuales el	café	semilla	está molido; la molienda uniforme produce una extracción r
, en la aplicación del programa del Consejo Veracruzano del	Café	semilla	, Covercafé, está sirviendo de enlace al menos durante el c
initario por una deuda que nunca terminó de saldar. </s><s>	Café	semilla	Tacvba engalanará despedida de fin de año </s><s> Este fi
n un promedio de 60 tazas por hora dependiendo del tipo de	café	infusión	a preparar. </s><s> Aníbal Prina fue liberado luego de que
cakes, yogur, mix de frutas o bowl de acai, acompañados de	café	infusión	de Loja, variedad de jugos. </s><s> Todos pueden halagar
afé", que es el encargado de compactar mediante presión el	café	semilla	molido. </s><s> Molienda automática. </s><s> Las nuevas
s, manzanas, naranjas, chocolate, huevos, fresas (frutillas) y	café	infusión	. </s><s> Después de consumir la más pequeña cantidad d
a la cafeína, y le encanta disfrutar de los que ofrecen en Mür	Café	establecimiento	, en la tranquila placita de Cristino Martos: "Es un sitio muy
. Alvorge. </s><s> Salimos del albergue y desayunamos en el	Café	establecimiento	Sancho I, donde habíamos comido ayer amablemente aten
/ el 100% toma más de dos tazas diarias (el 50% entre 2 y 3	café	infusión	, el 20% entre 3 y 4 y el 30% más de cuatro. </s><s> Pero,
vistas. </s><s> Las prácticas se realizan en negocios como	café	establecimiento	, tiendas, hoteles, restaurantes o negocios de comida rápid
mercado Viejo, un mercado ibicenco de verduras rodeado por	café	establecimiento	y bonitas tiendas. </s><s> La zona del puerto, La Marina, la
jella idea que nos emocionaba bastante. </s><s> Dos o tres	café	evento	después me interrumpió, esta vez con cara de susto: "¡No p
is vendidas en toda España, es potente y rápida, y realiza el	café	infusión	de la forma más rápida con sus 1000 watts de potencia, pa
ministro. </s><s> Una vez finalizado el ciclo de preinfusión, el	café	infusión	comienza a salir por la salida de café. 5 El suministro de ca

Corpus-based methodology

We used Sketch Engine (Kilgarriff et al., 2004) and the EsTenTen18 corpus.

We conducted manual corpus analysis using the annotation mode (Baisa et al., 2020).

We selected the nouns from a list of the most 5,000 frequent Spanish nouns taken from the EsTenTen and the Corpes XXI combined.

We first used groups of regular polysemy already found in the data (Renau et al., in preparation) (311 groups).

Annotation mode is active

ANNOTATION ACTIONS

MANAGE ANNOTATIONS

Left context

KWIC

Label

Right context

El sexto año se ha producido una fuerte caída del precio del **café** **semilla** y del azúcar, de modo que el Estado brasileño disminuye m

añana, y casi cada tarde: las medias

versar por una semana aceptó salir

i tarea y a finales de ese año nos lla

El Instituto Dominicano del Café. </s><s>

.. as, sesiones paralelas, puestos de i

paración de las dos superficies abra

, en la aplicación del programa del Cons

initario por una deuda que nunca terminó de saldar. </s><s> **café** **semilla** Tacyba engalanará despedida de fin de año </s><s> Este fi

n un promedio de 60 tazas por hora

cakes, yogur, mix de frutas o bowl d

afé", que es el encargado de compa

s, manzanas, naranjas, chocolate, h

a la cafeína, y le encanta disfrutar d

. Alvorge. </s><s> Salimos del alberg

/ el 100% toma más de dos tazas d

vistas. </s><s> Las prácticas se rea

mercado Viejo, un mercado ibicenco de verduras rodeado por **cafés** **establecimiento** y bonitas tiendas. </s><s> La zona del puerto, La Marina, la

jella idea que nos emocionaba bastante. </s><s> Dos o tres **cafés** **evento** después me interrumpió, esta vez con cara de susto: "¡No p

is vendidas en toda España, es potente y rápida, y realiza el **café** **infusión** de la forma más rápida con sus 1000 watts de potencia, pa

ministro. </s><s> Una vez finalizado el ciclo de preinfusión, el **café** **infusión** comienza a salir por la salida de café. 5 El suministro de ca

preparar café 'to prepare coffee' → beverage
sembrar café 'to plant coffee' → plant
café caliente 'hot coffee' → beverage
El café estaba lleno 'The café was full' → place

"Yo reaccioné de forma alérgica a los siguientes alimentos: plátanos, manzanas, naranjas, chocolate, huevos, fresas y **café**" 'I reacted allergic to the following foods: bananas, apples, oranges, chocolate, eggs, strawberries and coffee'.

Corpus-based methodology

We operate in a similar way of our previous project annotating verbs with the Corpus Pattern Analysis technique (Hanks, 2004, 2013).

However, part of the task has to do with pragmatic or "encyclopaedic" interpretation of the context.

We use a slightly modified version of the CPA Ontology for the selection of semantic types.

Preliminary results

252 analyzed nouns
28,972 analyzed
concordances

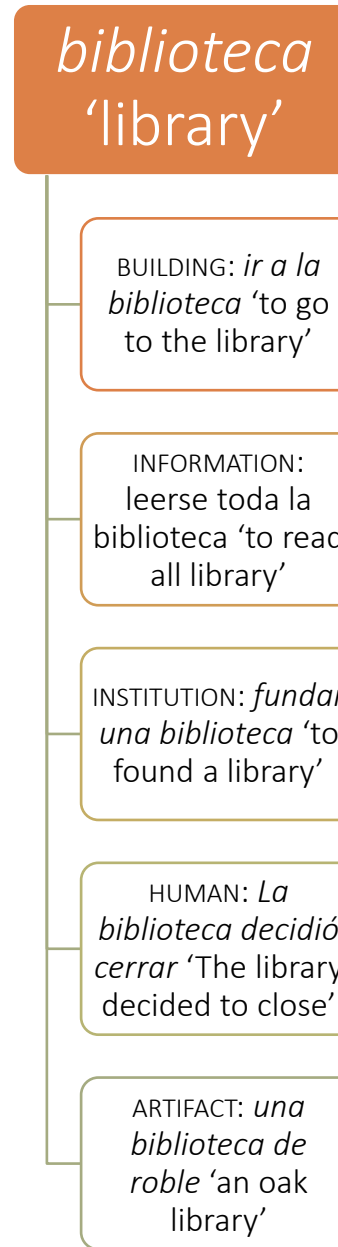
Main semantic types for group	Examples of analyzed nouns	Nr conc.	Examples of use
COLOUR / STUFF	<i>amatista, azabache, azurita, bronce, esmeralda</i>	3.032	vestido de amatista, extracción de amatista
MUSICAL INSTRUMENT OR VOICE / HUMAN	<i>acordeón, barítono, batería, contralto, guitarra</i>	5.175	tocar la guitarra, Las guitarras desafinaron
ACTION / RESULT	<i>agendamiento, alegato, apertrechamiento, apozamiento, austericidio</i>	3.750	durante su alegato, El film es una especie de alegato
ANIMAL / FOOD	<i>gallina, pollo, vaca</i>	450	criar pollos, comer pollo
ARTIFACT / INFORMATION / INSTITUTION / GENRE	<i>apunte, artículo, cuaderno, novela, revista</i>	2.500	manchar la novela, leer la novela, Lo contrataron en la revista, experto en novela decimonónica
INSTITUTION / POSITION / LOCATION / HUMAN / TIME PERIOD	<i>arzobispado, capitania, comisión, consulado, dirección</i>	1.844	organizar la presidencia, Fue promovido a la presidencia, explosión en la presidencia, La presidencia te ayudará, durante su presidencia
BUILDING / INSTITUTION / HUMAN / DIGITAL PLATFORM / EVENT	<i>aduana, ayuntamiento, banco, biblioteca, congreso</i>	4.219	La biblioteca está cerrada, el director de la biblioteca, La biblioteca rinde homenaje al escritor, Abre la biblioteca y descárgate los libros organizar un congreso
DANCE / MUSICAL COMPOSITION / GENRE / EVENT	<i>bachata, chachachá, cueca, guaguancó, milonga</i>	1.902	bailar una milonga, componer una milonga, experto en milonga, organizar una milonga
IDEOLOGY / HUMAN GROUP / TIME PERIOD / SYSTEM	<i>allendismo, anarquismo, comunismo, ecologismo, fascismo</i>	1.456	conceptos económicos del comunismo, crimen perpetrado por el comunismo, durante el comunismo, la caída del comunismo
PLANT / FLOWER / BEVERAGE / PART / COLOUR / LEAF / SEED / FRUIT / PRODUCTO / WOOD	<i>anís, café, mostaza, laurel, rosa</i>	2.150	plantar café, oler una rosa, Queremos dos cafés, falda mostaza, añadir laurel al guiso, untar la mostaza, mesa de laurel
CONTAINER / STUFF OR LIQUID / UNIT OF MEASURE	<i>acuario, balde, bañera, barril, biberón</i>	2.494	lavar las tazas, beberse una taza, añadir dos tazas de harina

Preliminary results

Pairs of semantic types are usually grouped in “chains” of more components.

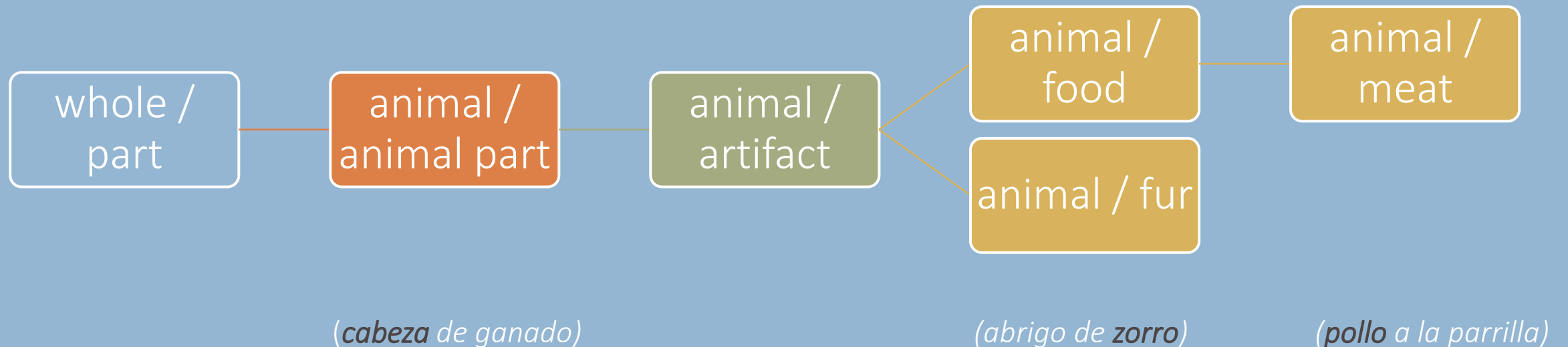
When a specific semantic type appears, it usually predicts the appearance of its pair in the corpus sample, e.g. HUMAN / INSTITUTION, INSTITUTION / BUILDING, ARTIFACT / INFORMATION.

Some of these pairs had not been reported in the literature.



Preliminary results

Types of regular polysemy can be classified in a hierarchy, from the more general to the more specific ones. E.g.:



Preliminary results

Many cases of **vagueness** in which both semantic types seem to be present, e.g.:

“Sacó su primer **libro** al año siguiente” ‘He/she got his first book out the following year’

“La poetisa publicó un solo **libro**” ‘The poet published only one book’

(ARTIFACT / INFORMATION)

Verbs which activate both meanings **simultaneously** could be *vender* ‘to sell’, *publicar* ‘to publish’, *comprar* ‘to buy’, etc.

Verbs for meaning **differentiation**: *escribir* ‘to write’ (for INFORMATION), *pesar* ‘to be heavy’ (for ARTIFACT), etc.

Rojo 'red'

Lema	Contexto izquierdo	Forma	Contexto derecho	Categoría gramatical	Anotación
rojo	es más común, donde incluso he necesitado aplicar medicamento en los folículos para poder controlarla, saludos </s><s>	roja	directa futbol en vivo gratis ver ... Disfrute y comparta lo mejor en partidos de futbol en vivo ... Y para poder ver la	S	Artefacto
rojo	S" en la capa en varios tamaños y colores. </s><s> A veces está en amarillo con delineado negro y otros artistas pusieron la "S" en	rojo	con un fondo amarillo. </s><s> La capa es un aspecto estilístico del traje, y más específicamente la forma en que se conecta al	S	Color
rojo	luego del pasillo tras la entrada aguardaba pacientemente con una sorpresa. </s><s> La sangre recorría la madera tiñendo de	rojo	su color opaco, al alzar la mirada del suelo se podían seguir varias líneas hasta llegar al punto alto de la pared, donde	S	Color
rojo	regionales. </s><s> A futuro vamos a plantear la opción de sacar las butacas y dejar sólo el cemento donde se podría pintar de	rojo	, así evitamos gastos", sostuvo el alcalde. </s><s> Nos visitará la arquitecta María José Leveratto quien nos hablará de la	S	Color
rojo	, las diferencias políticas se desarrollan en un saludable clima de intercambio de ideas, nadie cruza el semáforo en	rojo	, los motociclistas no roban, los supermercados son pletóricos y la gente anda por la calle con el teléfono en la mano. </s>	S	Color
rojo	mundo"), donde se indica con diferentes colores el ratio de infección –desde el verde, para regiones seguras, hasta el	rojo	oscuro, para brotes severos de malware– Este mapa, además, es interactivo: "planeando" con el cursor sobre los	S	Color
rojo	" delanteros, firmados por Peugeot Sport y tapizados en combinación de Alcantara y una malla negra con pespunte en	rojo	, y en la decoración en negro lacado del habitáculo –también con contrastes en rojo en los cinturones o las alfombrillas	S	Color
rojo	hecho a los que ofrece el propio Juke. </s><s> Mientras que en la gama del Juke los tonos disponibles son amarillo, blanco, azul,	rojo	, gris y negro, para el Nissan Note se marca la diferencia con algunas tonalidades nuevas. </s><s> De esta forma, en el Nissan	S	Color
rojo	colegiado también evitaba una goleada escandalosa al mirar para otro lado en un claro penalti a Airam y no sancionar con	roja	una agresión al canario. </s><s> El ariete de Puerto de la Cruz relegaba a un Villar que se marchaba lesionado, el único lunar de	S	Artefacto
rojo	cosechado en el Bernabeu, perdida la Copa ante el Real Madrid, los de Guardiola deben centrarse en el encuentro ante los	rojillos	para mantener su cómoda ventaja en el campeonato (ocho puntos) y restar un partido al calendario. </s><s> Real-Valencia. </s><s> El	S	Persona
rojo	y la región todavía se mantiene en latín el nombre de la provincia (Estiria) y el escudo común de los brazos de plata,	rojo	y armados con cuernos. </s><s> Por lo tanto, hay una distinción entre el concepto geográfico de Estiria y el oficial-político	S	Color
rojo	en forma de corazón para expresar un claro mensaje de amor: I love you. </s><s> El osito es de tonos marrones combinando con	rojo	y marrón más claro. </s><s> Demuéstrale tu amor con este encantador osito de peluche. </s><s> Tamaño 25 cm. </s><s> "Fue una sorpresa enorme, no	S	Color
rojo	López de Uralde, reaccionó con flema en su cuenta de Twitter: "Atacan la sede de @equo en Madrid dejando pintadas de "	Rojos	No". </s><s> No se enteran: somos verdes". </s><s> En el mismo distrito de Madrid el Partido Comunista recibió la visita de unos	S	Persona
rojo	como el club ganó su primera Copa de Alemania en 1968. </s><s> Incluso en 1974 el equipo fue relegado a la segunda división, los	rojios	no se dieron por vencidos y en 2008 puso en marcha un proyecto para traer de nuevo la vieja gloria del club. </s><s> En 2010 el	S	Persona
rojo	azul oscuro apenas presentan diferencias y en ocasiones resultan difíciles de distinguir. </s><s> Combinaciones de verde y	rojo	suelen ser desagradables a la vista. </s><s> Otras combinaciones que debes evitar son rojo-azul, verde-azul y marrón-negro. </s>	S	Color
rojo	que las paredes, los techos y los pisos estén revestidos o pintados en este maravilloso color y que luego, los salones en	rojo	rubí estén equipados con muebles en color marrón, negro o blanco. </s><s> Esta combinación se ve fabulosa. </s><s> Otra opción sería	S	Color
rojo	final, los goles de Mallarach, Víctor Gutiérrez y Roger Tahull en un parcial espectacular de 3–0 han dejado el choque al	rojo	vivo. </s><s> Grecia ha logrado adelantarse a 1:20 en superioridad y Munárriz ha tenido un último remate que se ha estrellado	S	Color
rojo	has conseguido traer todo esto aquí? </s><s> - No tranquilo no me iba a reir pero es que tu edredón es ""muy de Gryffindor"" con ese	rojo	fuego y el león dorado. - le dije riendo. - Así que te has estado leyendo los siete libros que úsamos en el colegio. </s><s> ¿Me lo	S	Color
rojo	que si es cierto esque uefa no usa misma vara de medir, a pepe por una jugada parecida a esta en la que si toca balon le saquen	roja	, y al del dinamo no le saque ni amarilla por una cuchillada. </s><s> Por lo demas, decir que benzema es un jugador de un talento	S	Artefacto

Colour								
<i>Rojo</i>			<i>Amarillo</i>			<i>Azul</i>		
Tipo semántico	n	%	Tipo semántico	n	%	Tipo semántico	n	%
Artefacto	11	8.40	Artefacto	39	21.31			
Color	93	70.99	Color	131	71.58	Color	130	87.25
Institución	11	8.40	Persona	12	6.56	Institución	3	2.01
Persona	16	12.21	Pez	1	0.55	Persona	16	10.74
Total sustantivos	131	100	Total sustantivos	183	100	Total sustantivos	149	100

Colour								
<i>Verde</i>			<i>Naranja</i>			<i>Violeta</i>		
Tipo semántico	n	%	Tipo semántico	n	%	Tipo semántico	n	%
			Color	99	23.19			
Color	69	69.00	Fruta	269	63.00	Color	48	57.14
Institución	9	9.00	Institución	4	0.94	Flor	14	16.67
Persona	14	14.00	Persona	11	2.58	Planta	9	10.71
Vegetación	8	8.00	Sabor/aroma	44	10.30	Sabor/aroma	13	15.48
Total sustantivos	100	100	Total sustantivos	427	100	Total sustantivos	84	100

Final remarks

Corpus analysis provides with richer and complex information than the one that we have in the literature or Spanish dictionaries.

For the moment, regular polysemy was found in all analysed words. It seems to be a **wider phenomenon** than expected.

We still have to add more nouns to the analysis, and create a **taxonomy** of types of regular polysemy.

Lexicographic representation of regular polysemy could include, among other things, the use of **signposts** with a general description of the semantic type. Some of the meanings may not require a specific definition, only corpus examples.

Thank you!

irene.renau@pucv.cl

www.tecling.com

