

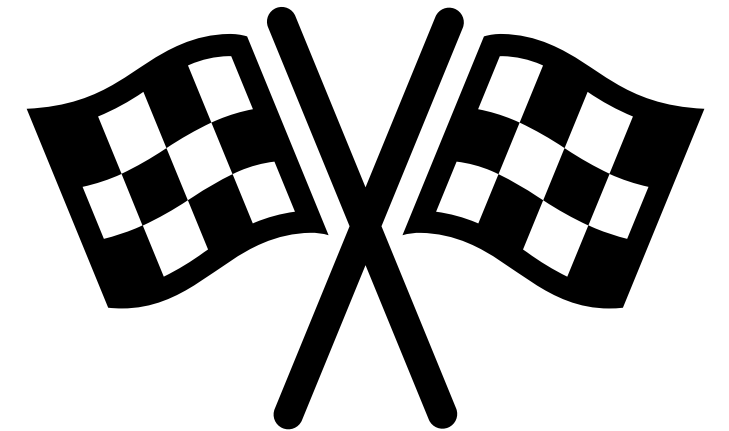
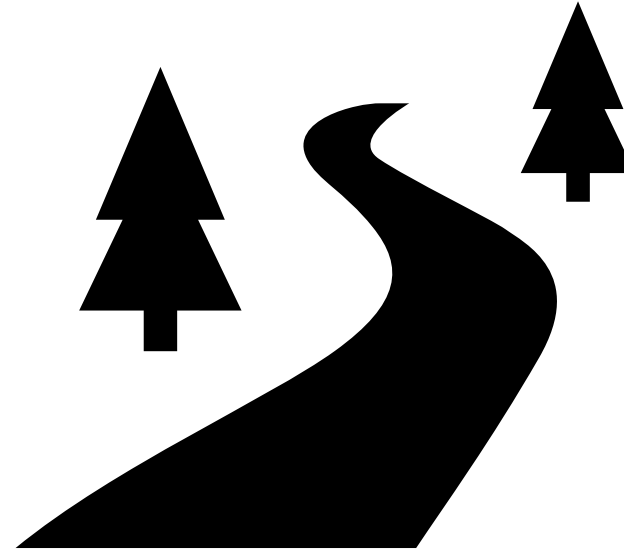
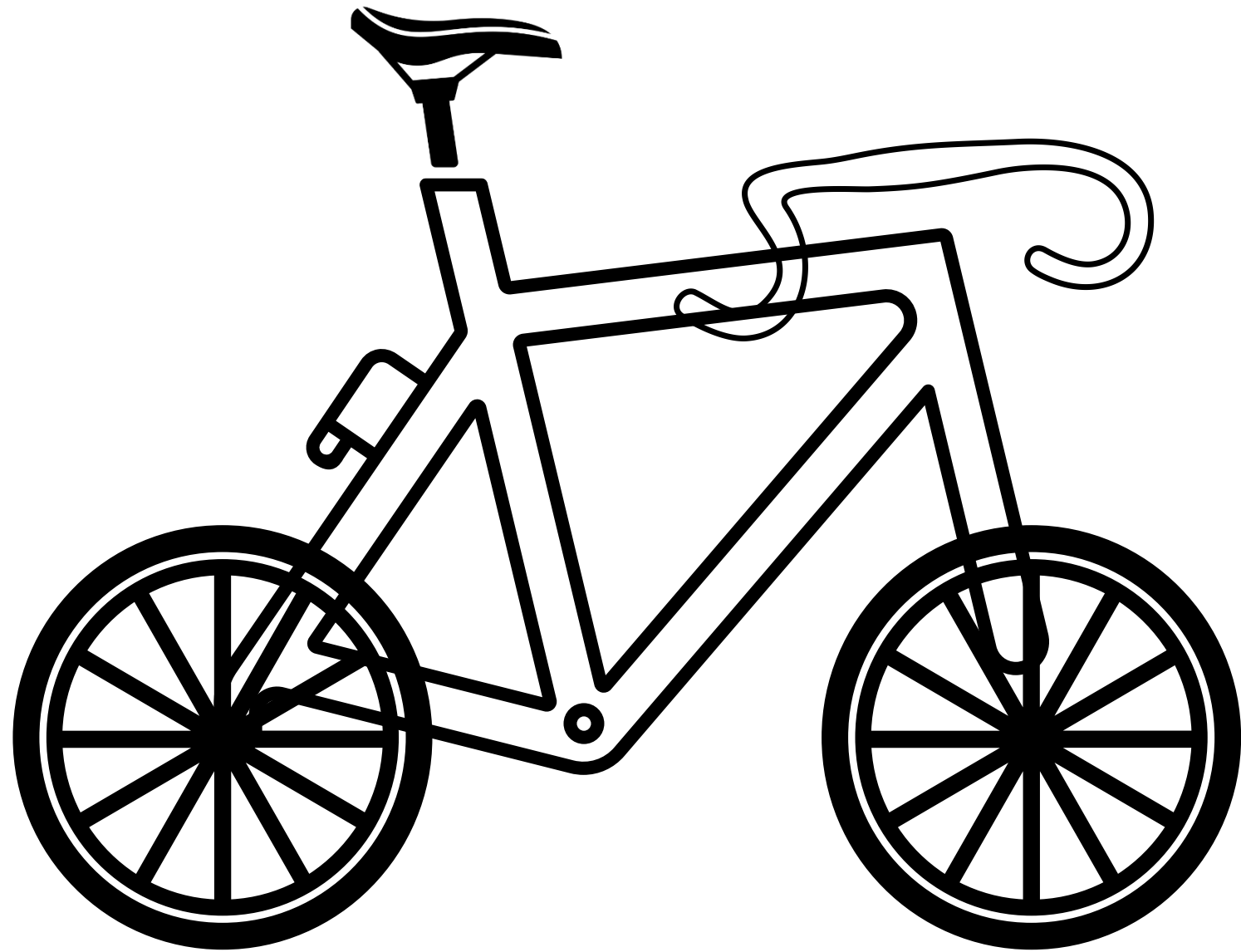
CORPUS-BASED EXAMPLE SENTENCE

SELECTION FOR L2 SPANISH: FROM

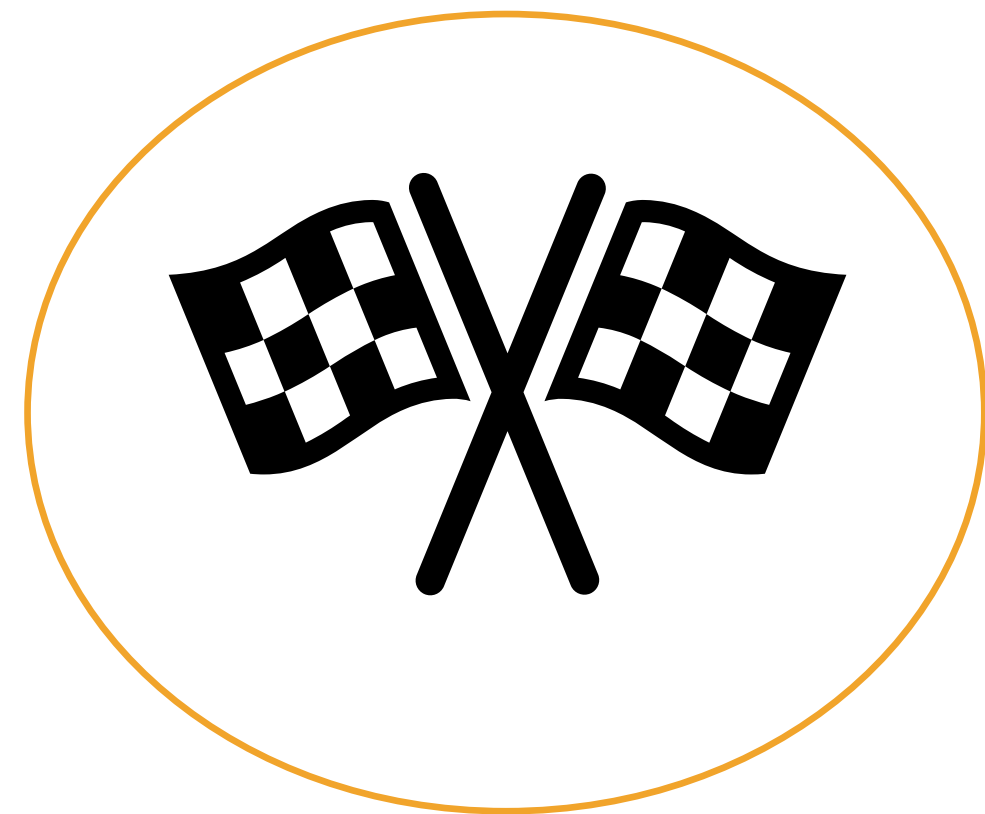
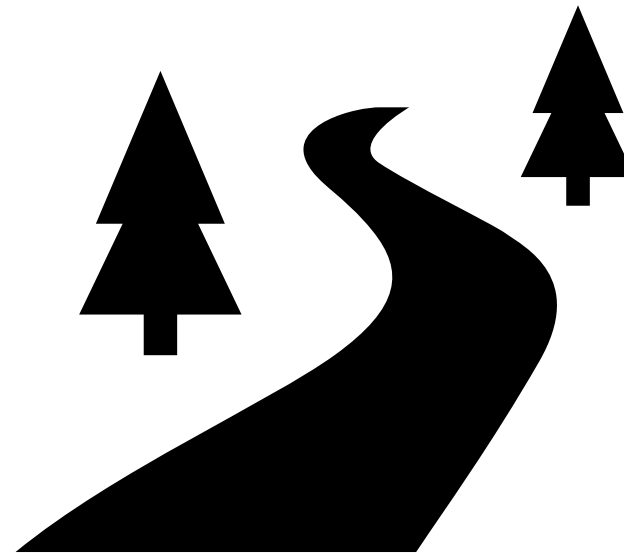
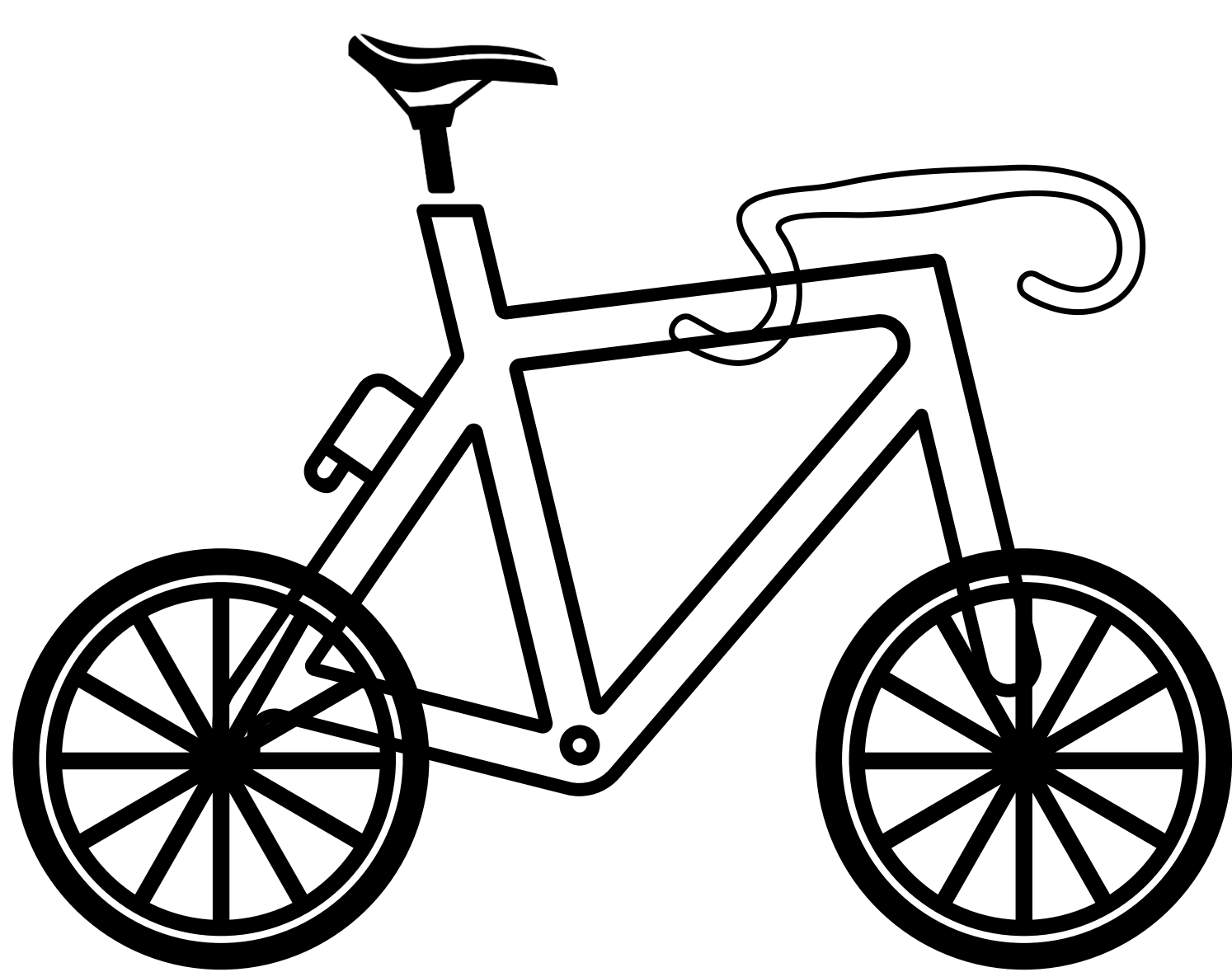
FRAMEWORK TO PRACTICE

Jasper Degraeuwe – TaLC 2024 – 10 June 2024

LET'S GO FOR A RIDE

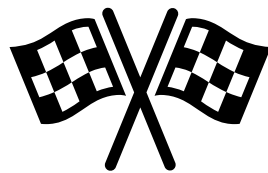


GOAL: LANGUAGE LEARNING MATERIALS



GOAL: LANGUAGE LEARNING MATERIALS

Vocabulary item	POS
Creative	ADJ
Edit	VERB
Font	NOUN
Infographic	NOUN
Logo	NOUN
Typography	NOUN
User experience design	NOUN



GOAL: LANGUAGE LEARNING MATERIALS

Vocabulary item	POS	Example sentence
Creative	ADJ	An interesting view can stimulate creative thinking.
Edit	VERB	The recorded interview was edited for broadcast.
Font	NOUN	The font size should be 12 pt.
Infographic	NOUN	Below is an infographic that highlights clinic activities.
Logo	NOUN	The finely detailed interior woodwork also includes company logos .
Typography	NOUN	Computer technology revolutionized typography in the 20th century.
User experience design	NOUN	Finally, there are user interface and user experience design issues that must be addressed.



banco

bank; bench



banco

banco

Voy al **banco** para abrir una cuenta.

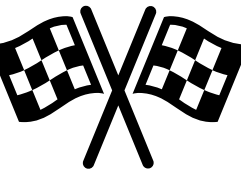
bank; bench

bank

banco

Al llegar al parque, me senté en un **banco**.

bench



Unit 3: The passive voice

In Spanish, the passive voice can be expressed in two different ways:

1. The verb *ser* plus the past participle of the lexical verb. The object of the action becomes the subject of the sentence and there is concordance between the number and gender of the subject and the past participle. The agent of the action is introduced by the preposition *por* (but can be left out if desired).
2. A so-called *pasiva refleja* with the pronoun *se* followed by the lexical verb. The object of the action again is the subject of the sentence and there is concordance between the number of the subject and the verb. The agent of the action cannot be added.



Unit 3: The passive voice

In Spanish, the passive voice can be expressed in two different ways:

1. The verb *ser* plus the past participle of the lexical verb. The object of the action becomes the subject of the sentence and there is concordance between the number and gender of the subject and the past participle. The agent of the action is introduced by the preposition *por* (but can be left out if desired).

Las noticias fueron publicadas por el periódico. ('The news items were published by the newspaper.')
--

La conferencia es organizada por un grupo de académicos muy apasionados. ('The conference is organised by a group of very passionate scholars.')
--

2. A so-called *pasiva refleja* with the pronoun *se* followed by the lexical verb. The object of the action again is the subject of the sentence and there is concordance between the number of the subject and the verb. The agent of the action cannot be added.

Las noticias se publicaron.

La conferencia se organiza en formato híbrido. ('The conference is organised in a hybrid format.')
--



INTERMEDIATE CONCLUSION

Example sentences are important

GOAL: LANGUAGE LEARNING MATERIALS

Rellena los huecos

(Aviso: las traducciones provienen de un glosario descontextualizado y dan la forma no conjugada. Véase la clave al final del documento.)

lemas : consumidor, innovación, medioambiente, proveedor, sostenibilidad

1. BIC adapta y desarrolla sus productos de acuerdo con las cambiantes necesidades de los c..... . *(Glosario: consument, gebruiker, verbruiker)*
2. Las materias primas de los p..... deben cumplir unos requisitos muy precisos y superar controles de calidad muy estrictos . *(Glosario: leverancier)*
3. Generación tras generación , la marca Dr. Oetker ha ofrecido a sus clientes y c.....
..... productos que garantizan el éxito con un sabor excelente y de máxima calidad .
(Glosario: consument, gebruiker, verbruiker)
4. La calidad , el m..... , la prevención y la seguridad son elementos que forman parte integral de las actividades y cultura en Applus+ . *(Glosario: milieu)*
5. El primero de nuestros diez objetivos de s..... es “ una vida segura , saludable y de calidad para nuestros empleados ” . *(Glosario: duurzaamheid)*



GOAL: LANGUAGE LEARNING MATERIALS

Traduce al inglés

1. BIC adapta y desarrolla sus productos de acuerdo con las cambiantes necesidades de los consumidores.

.....

.....

.....

2. Las materias primas de los proveedores deben cumplir unos requisitos muy precisos y superar controles de calidad muy estrictos .

.....

.....

.....

3. La calidad , el medioambiente , la prevención y la seguridad son elementos que forman parte integral de las actividades y cultura en Applus+ .

.....

.....

.....



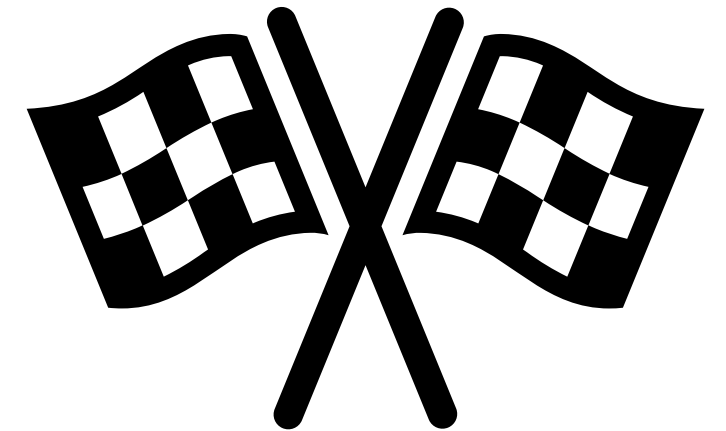
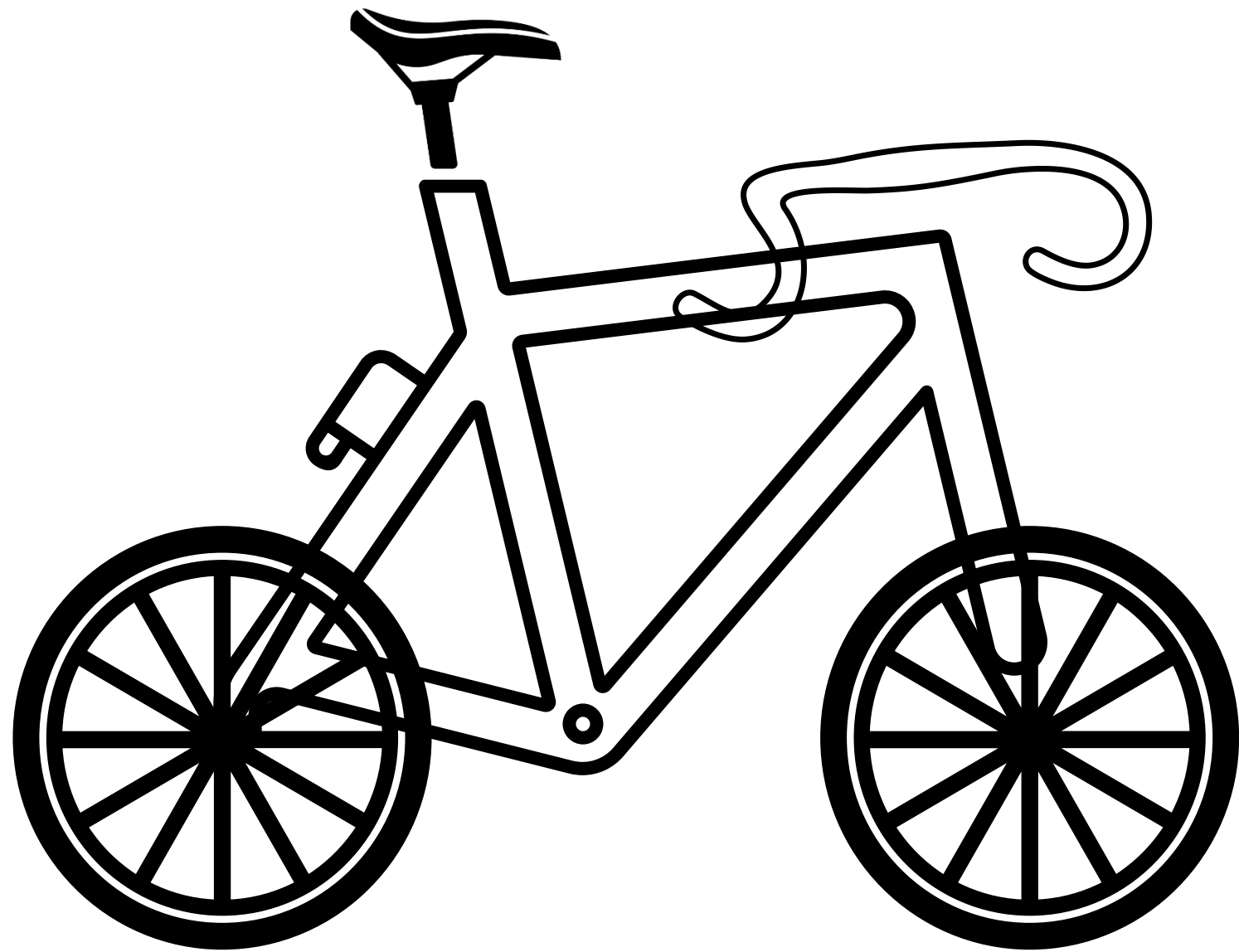
FULL CONCLUSION

- **Need for “good” example sentences and/or authentic contexts**

FULL CONCLUSION

- **Need for “good” example sentences and/or authentic contexts**
- Strengthens word form-word meaning associations (Laufer & Shmueli, 1997)
- Fosters both language comprehension and production (Frankenberg-Garcia, 2014)

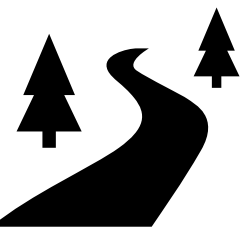
LET'S GO FOR A RIDE



RELATED RESEARCH

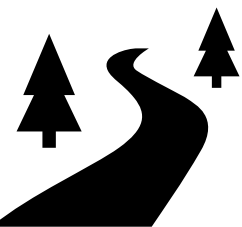
METHODS TO OBTAIN EXAMPLE SENTENCES

1. Intuited / invented examples



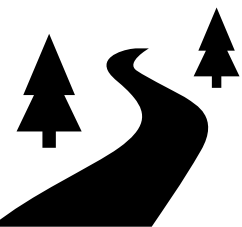
METHODS TO OBTAIN EXAMPLE SENTENCES

1. Intuited / invented examples
2. Corpus-based example selection



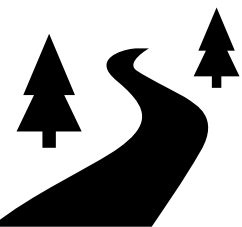
CORPUS-BASED EXAMPLES

- Towards end of 1980s
- Rise of online accessible corpora
- Advances in corpus query tools
- With posterior human revision or without (COBUILD)
- Specific selection algorithms (e.g., GDEX; Kilgarrriff et al., 2008)



METHODS TO OBTAIN EXAMPLE SENTENCES

1. Intuited / invented examples
2. Corpus-based example selection
3. **Generative Artificial Intelligence**



GenAI-BASED EXAMPLES



You

Give me three example sentences of no longer than 20 words that clarify the meaning and usage of the word "ubiquitous".



ChatGPT

1. Smartphones are now ubiquitous, found everywhere from bustling cities to remote villages.
2. The internet has made information ubiquitous, accessible to anyone with a connection.
3. Coffee shops are ubiquitous in urban areas, appearing on nearly every street corner.



GenAI-BASED EXAMPLES



You

ubiquitous → Sentence 1: "Smartphones are now ubiquitous, found everywhere from bustling cities to remote villages." | Sentence 2: "The internet has made information ubiquitous, accessible to anyone with a connection." | Sentence 3: "Coffee shops are ubiquitous in urban areas, appearing on nearly every street corner."

nonplussed →



ChatGPT

Sentence 1: "Despite her best efforts, Jane was nonplussed by the unexpected turn of events, unsure of how to react." | Sentence 2: "The politician remained nonplussed by the reporter's probing questions, maintaining a calm demeanor throughout the interview." | Sentence 3: "His nonplussed expression betrayed his confusion at the complex instructions given by his boss."



GenAI-BASED EXAMPLES



You

ubiquitous → Sentence 1: "Smartphones are now ubiquitous, found everywhere from bustling cities to remote villages." | Sentence 2: "The internet has made information ubiquitous, accessible to anyone with a connection." | Sentence 3: "Coffee shops are ubiquitous in urban areas, appearing on nearly every street corner."

nonplussed →



ChatGPT

Sentence 1: "Despite her best efforts, Jane was nonplussed by the unexpected turn of events, unsure of how to react." | Sentence 2: "The politician remained nonplussed by the reporter's probing questions, maintaining a calm demeanor throughout the interview." | Sentence 3: "His nonplussed expression betrayed his confusion at the complex instructions given by his boss."



Few-shot prompting

GenAI-BASED EXAMPLES

JA

You

Give me three example sentences of no longer than 20 words that clarify the meaning and usage of the word "ubiquitous".

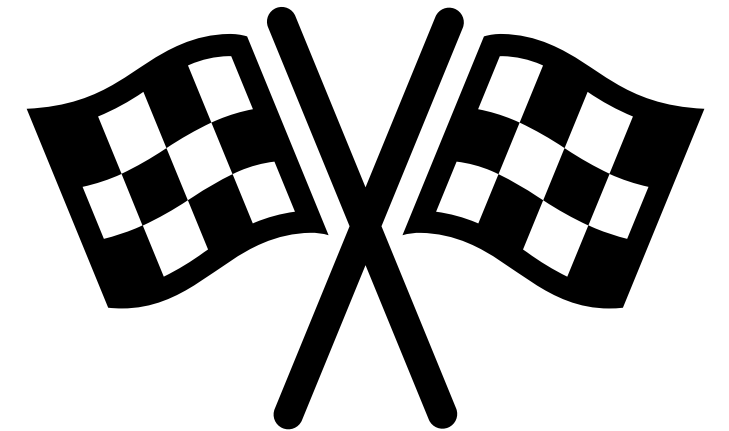
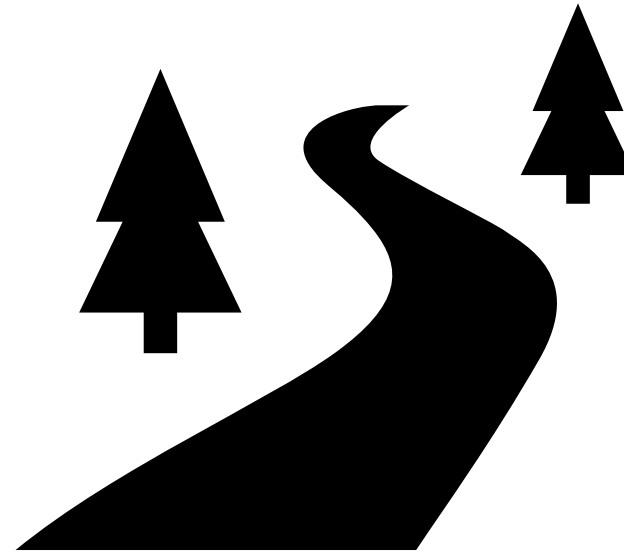
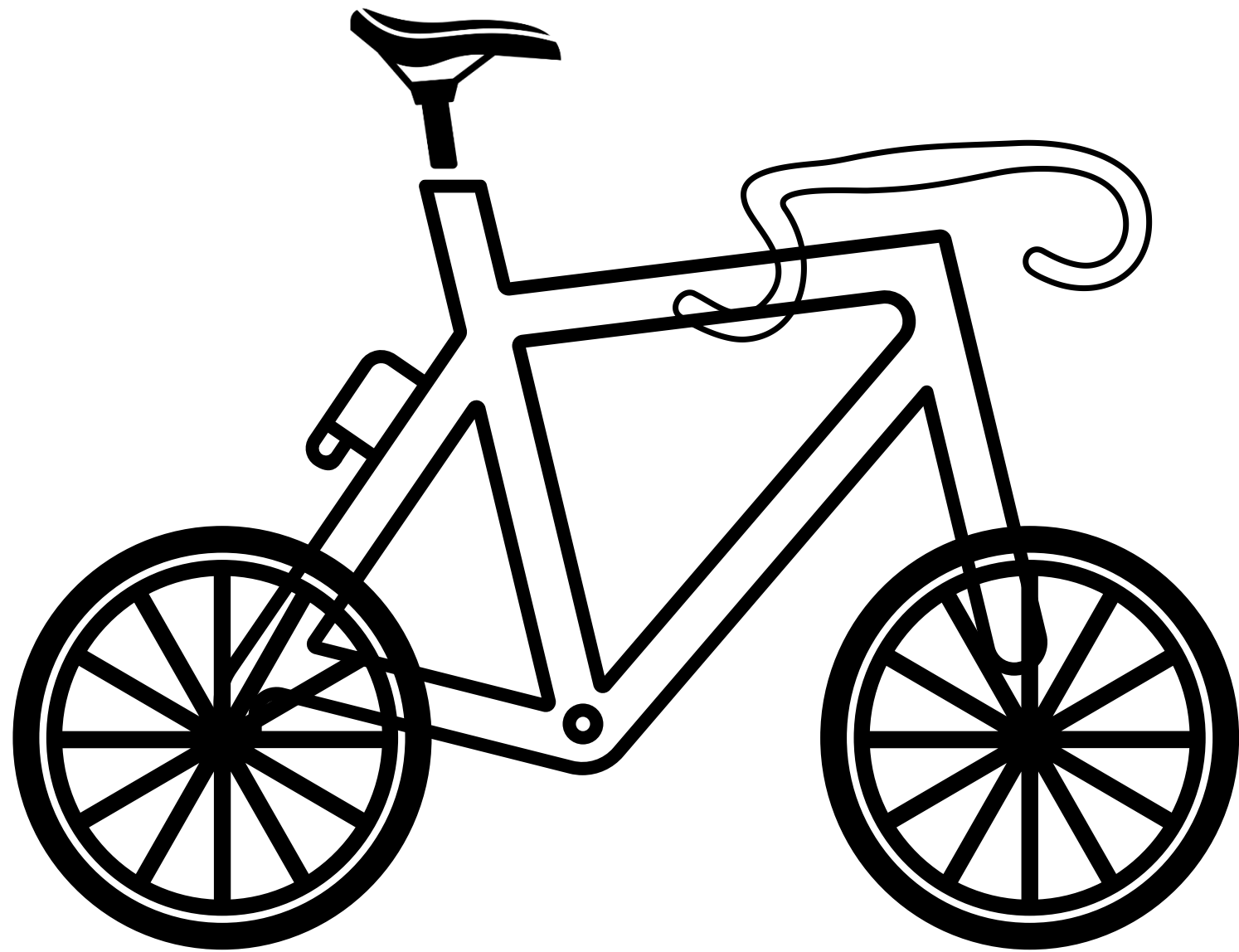


ChatGPT

1. Smartphones are now ubiquitous, found everywhere from bustling cities to remote villages.
2. The internet has made information ubiquitous, accessible to anyone with a connection.
3. Coffee shops are ubiquitous in urban areas, appearing on nearly every street corner.

Zero-shot prompting

LET'S GO FOR A RIDE





RESEARCH GAPS

- Frameworks to select **pedagogically suitable** example sentences from corpora
- Evaluate pedagogical usability and validity of **GenAI-based example sentences**
- Languages other than English

METHODOLOGY

METHODOLOGY

1. Build framework to select examples from corpora for L2 Spanish
2. Define set of target vocabulary items
3. Use framework to select pedagogically suitable examples for those items
4. Use GenAI model to create examples for those items
5. Compare both types in experiment with L2 learners

METHODOLOGY

1. Build framework to select examples from corpora for L2 Spanish
2. Define set of target vocabulary items
3. Use framework to select pedagogically suitable examples for those items
4. Use GenAI model to create examples for those items
5. Compare both types in experiment with L2 learners

FRAMEWORK FOR L2 SPANISH

- Foundational work: HitEx for L2 Swedish (Pilán et al., 2016)

HitEx (PILÁN ET AL., 2016)


Nr	Criterion	Nr	Criterion
	Search term		Additional structural criteria
1	<i>Absence of search term</i>	13	Negative formulations
2	Number of matches	14	<i>Interrogative sentence</i>
3	<i>Position of search term</i>	15	<i>Direct speech</i>
	Well-formedness	16	<i>Answer to closed questions</i>
4	<i>Dependency root</i>	17	Modal verbs
5	Ellipsis	18	Sentence length
6	<i>Incompleteness</i>		Additional lexical criteria
7	Non-lemmatized tokens	19	Difficult vocabulary
8	Non-alphabetical tokens	20	Word frequency
	Context independence	21	Out-of-vocabulary words
9	<i>Structural connective in isolation</i>	22	Sensitive vocabulary
10	Pronominal anaphora	23	Typicality
11	Adverbial anaphora	24	Proper names
12	L2 complexity in CEFR level	25	Abbreviations

FRAMEWORK FOR L2 SPANISH

- Foundational work: HitEx for L2 Swedish
 - Our work
 - Convert HitEx into Spanish (e.g., take into account pronoun dropping)
 - Expand (personalised difficulty classifier; more importance attributed to *typicality* criterion)
- **SeEjemCor** framework (**Selección de Ejemplos de Corpus**)

Nr	Criterion
/	Proficiency level target audience.
/	Number of years experience target audience.
1	Boolean value indicating if search term has to occur in sentence.
2	Numerical value indicating maximum number of times search term can occur in sentence.
3	Numerical value between 0 and 1 indicating at which position search term has to occur.
4	Boolean value indicating if sentence has to contain dependency root.
5	Boolean value indicating if sentence has to contain subject or finite verb.
6	Boolean value indicating if sentence has to contain explicit subject.
7	Boolean value indicating if sentence has to start with capital letter and end with punctuation mark.
8	Numerical value indicating maximum number of tokens which do not occur in SCAP-based lemma lexicon.
9	Numerical value indicating maximum number of non-alphabetical tokens (e.g., mark-up traces in web materials).
10	Boolean value indicating that no conjunction or subjunction can appear in sentence-initial position.
11	Numerical value indicating maximum number of demonstrative pronouns (e.g., <i>este esta</i> : ‘this’; <i>ese esa</i> : ‘that’).
12	Numerical value indicating maximum number of words/phrases which occur in precompiled list of anaphoric expressions (e.g., <i>allí</i> : ‘there’; <i>aquí</i> : ‘here’; <i>entonces</i> : ‘then’).
13	Numerical value indicating maximum number of negation adverbials (e.g., <i>no</i> : ‘no’; <i>nadie</i> : ‘nobody’; <i>nada</i> : ‘nothing’).
14	Boolean value indicating that sentence cannot represent direct question.

- 15 Boolean value indicating that sentence cannot represent direct speech (i.e. speaking verb combined with delimiters such as quotation marks).
- 16 Boolean value indicating that sentence cannot represent answer to closed question (i.e. sentence-initial adverb of affirmation or negation followed by delimiter).
- 17 Numerical value indicating maximum number of tokens which occur in precompiled list of modal verbs (when functioning as an auxiliary verb).
- 18 Numerical value indicating maximum number of tokens in the sentence (including punctuation).
- 19 Numerical value indicating maximum number of words above the proficiency level of the target audience according to a personalised machine learning classifier.
- 20 Numerical value indicating minimum frequency of words in SCAP lemma frequency dictionary (expressed in percentiles).
- 21 Numerical value indicating maximum number of words not included in SCAP token lexicon.
- 22 Boolean value indicating that sentence cannot contain words which occur in precompiled list of potentially sensitive words related to PARSNIP topics.
- 23 Numerical value indicating maximum number of proper names.
- 24 Numerical value indicating minimum average normalised Lexicographer's Mutual Information (Bouma, 2009) score for verb-noun pairs (in subject, object, and oblique relation) and all noun-adjective pairs (in attributive or predicative relation) in the sentence. The scores are retrieved from a SCAP-based resource.
- 25 Numerical value indicating minimum average ΔP score (Ellis, 2006; Gries, 2013) for verb-noun pairs (in subject, object, and oblique relation) and all noun-adjective pairs (in attributive or predicative relation) that include the search term. The scores are retrieved from a SCAP-based resource.
- 26 Numerical value indicating minimum average cosine similarity of search term with head and dependents (both static and contextualised word embeddings).
- 27 Numerical value indicating minimum average n -gram frequency of the sentence (excluding n -grams with punctuation marks). Frequencies are retrieved from SCAP dictionary containing lemma-based n -grams.



Product ▾ Solutions ▾ Resources ▾ Open Source ▾ Enterprise ▾ Pricing

Search

JasperD-UGent / example-sentence-selector Public

<> Code

Issues

Pull requests

Actions

Projects


Security

Insights

main ▾ 1 Branch 0 Tags

Go to file

<> Code ▾

 JasperD-UGent

Update README.md

8f2384e · 3 weeks ago

🕒 14 Commits

input	Delete input/datasets_raw/custom_preprocessed/resExp dire...	3 weeks ago
utils	Upload scripts	3 weeks ago
.gitattributes	Initial commit	last year
.gitignore	Create .gitignore	last year
LICENSE	Update LICENSE	3 months ago
README.md	Update README.md	3 weeks ago
copyrightNotice.txt	Update copyrightNotice.txt	3 months ago
exampleSentenceSelector.py	Upload scripts	3 weeks ago
exampleSentenceSelector_defs.py	Upload scripts	3 weeks ago



METHODOLOGY

1. Build framework to select examples from corpora for L2 Spanish
- 2. Define set of target vocabulary items**
3. Use framework to select pedagogically suitable examples for those items
4. Use GenAI model to create examples for those items
5. Compare both types in experiment with L2 learners

DEFINE SET OF TARGET ITEMS

- 250 items for beginner learners
- 250 items for advanced learners

DEFINE SET OF TARGET ITEMS

- **250 items for beginner learners**
 - **1,001 – 2,000 range in Davies and Hayward**
Davies (2018) frequency list
 - **150 NOUN, 50 VERB, 50 ADJ**
- 250 items for advanced learners

DEFINE SET OF TARGET ITEMS

- 250 items for beginner learners
- **250 items for advanced learners**
 - **25M domain-specific corpus containing newspaper articles on legal topics ranked based on keyness (Odds Ratio effect size metric)**
 - **150 NOUN, 50 VERB, 50 ADJ**

DEFINE SET OF TARGET ITEMS

- 250 items for beginner learners
- 250 items for advanced learners

DEFINE SET OF TARGET ITEMS

- 250 items for beginner

Beginner

escritor | NOUN
(1011)
riesgo | NOUN
(1016)
plaza | NOUN
(1020)
enemigo | NOUN
(1024)
raíz | NOUN
(1028)

intento | NOUN
(1012)
policía | NOUN
(1017)
llamada | NOUN
(1022)
unidad | NOUN
(1025)
compromiso | NOUN
(1029)

regla | NOUN
(1014)
estudiante | NOUN
(1019)
sensación | NOUN
(1023)
idioma | NOUN
(1027)
comentario | NOUN
(1030)

DEFINE SET OF TARGET ITEMS

for beginners

Beginner

regla | NOUN
(1014)

estudiante | NOUN
(1019)

asación | NOUN

NOUN

Advanced

cohecho | NOUN
(1796.97)

lla |
(1022)

unidad | NOUN
(1025)

compromiso | NOUN
(1029)

adjudicación | NOUN
(1665.37)

recusación | NOUN
(2034.42)

escritor | NOUN
(1011)

riesgo | NOUN
(1016)

plaza | NOUN
(1020)

enemigo | NOUN
(1024)

raíz | NOUN
(1028)

METHODOLOGY

1. Build framework to select examples from corpora for L2 Spanish
2. Define set of target vocabulary items
- 3. Use framework to select pedagogically suitable examples for those items**
4. Use GenAI model to create examples for those items
5. Compare both types in experiment with L2 learners

SELECT CORPUS SENTENCES

- Beginner group: 7.5M corpus containing accessible reportages about tourist destinations
- Advanced group: 25M domain-specific corpus containing newspaper articles on legal topics
- Select top-ranked sentence according to framework

SELECT CORPUS SENTENCES: ILLUSTRATION

Target vocabulary item = *complaint*

ID	Rank	Sentence
1	?	When asked if I had any complaints about the TaLC conference, I could only say that
2	?	Investigators said 170 complaints were made, with 39 of the complaints resulting in prosecutions.
3	?	She attended the local police station to file a complaint after being denied access to her granddaughter.

SELECT CORPUS SENTENCES: ILLUSTRATION

ID	Rank	1	2	3	...	7	...	25	26	27
1	?	True	1	0.24	...	False				
2	?	True	2	0.11	...	True				
3	?	True	1	0.5	...	True				

Nr	Criterion	V _{set} BEG	V _{set} ADV
/	Proficiency level target audience.	B1	C1
/	Number of years experience target audience.	1	3
1	Boolean value indicating if search term has to occur in sentence.	<i>True</i>	<i>True</i>
2	Numerical value indicating maximum number of times search term can occur in sentence.	1	1
3	Numerical value between 0 and 1 indicating at which position search term has to occur.	<i>None</i>	<i>None</i>
4	Boolean value indicating if sentence has to contain dependency root.	<i>True</i>	<i>True</i>
5	Boolean value indicating if sentence has to contain subject or finite verb.	<i>True</i>	<i>True</i>
6	Boolean value indicating if sentence has to contain explicit subject.	<i>True</i>	<i>True</i>
7	Boolean value indicating if sentence has to start with capital letter and end with punctuation mark.	<i>True</i>	<i>True</i>

SELECT CORPUS SENTENCES: ILLUSTRATION

ID	Rank	1	2	3	...	7	...	25	26	27
1	?	True	1		...	False		0.1	0.5	0
2	?	True	2		...	True		0.05	0.61	18
3	?	True	1		...	True		0.32	0.58	85.3

25	Numerical value indicating minimum average ΔP score (Ellis, 2006; Gries, 2013) for verb-noun pairs (in subject, object, and oblique relation) and all noun-adjective pairs (in attributive or predicative relation) that include the search term. The scores are retrieved from a SCAP-based resource.	<i>all</i>	<i>all</i>
26	Numerical value indicating minimum average cosine similarity of search term with head and dependents (both static and contextualised word embeddings).	<i>all</i>	<i>all</i>
27	Numerical value indicating minimum average n -gram frequency of the sentence (excluding n -grams with punctuation marks). Frequencies are retrieved from SCAP dictionary containing lemma-based n -grams.	<i>all</i>	<i>all</i>

SELECT CORPUS SENTENCES: ILLUSTRATION

ID	Rank	1	2	3	...	7	...	25	26	27
1	EXCL	True	1		...	False		0.1	0.5	0
2	2	True	2		...	True		0.05	0.61	18
3	1	True	1		...	True		0.32	0.58	85.3

METHODOLOGY

1. Build framework to select examples from corpora for L2 Spanish
2. Define set of target vocabulary items
3. Use framework to select pedagogically suitable examples for those items
- 4. Use GenAI model to create examples for those items**
5. Compare both types in experiment with L2 learners

CREATE GenAI-BASED SENTENCES

- Different system prompts
- Different user prompts

CREATE GenAI-BASED SENTENCES

- Zero shot
 - General
 - General + target audience
 - General + target audience + criteria
- Few-shot

CREATE GenAI-BASED SENTENCES

- Zero-shot, general
 - **System prompt:** *You are a teacher of Spanish as a foreign language.*
 - **User prompt:** *Write a sentence between 15 and 30 words in Spanish that presents an authentic usage of the Spanish [POS] '[WORD]', a vocabulary item that has to be learnt by your students. The sentence should not be a definition of the word.*

CREATE GenAI-BASED SENTENCES

- Zero-shot, general + target audience
 - **System prompt (beginner):** *You are a teacher of Spanish as a foreign language to a beginner/lower-intermediate group of university students who have been studying Spanish for one year.*
 - **System prompt (advanced):** *You are a teacher of Spanish as a foreign language to an upper-intermediate/advanced group of university students who have been studying Spanish for three years.*
 - **User prompt:** *Write a sentence between 15 and 30 words in Spanish that presents an authentic usage of the Spanish [POS] '[WORD]', a vocabulary item that has to be learnt by your students. The sentence should not be a definition of the word.*

CREATE GenAI-BASED SENTENCES

Subset	POS	Prompt type
1	30 NOUN 10 VERB 10 ADJ	General
2		General + target audience
3		General + target audience + criteria
4		Few shot
5		<ul style="list-style-type: none">• General• General + target audience• General + target audience + criteria• Few shot

METHODOLOGY

1. Build framework to select examples from corpora for L2 Spanish
2. Define set of target vocabulary items
3. Use framework to select pedagogically suitable examples for those items
4. Use GenAI model to create examples for those items
5. Compare both types in experiment with L2 learners

CORPUS-BASED VERSUS Gen-AI

- Beginner group: 3 \approx B1 Spanish L2 learners (L1 = Dutch)
- Advanced group: 4 \approx C1 Spanish L2 learners (L1 = Dutch)
- Evaluate all subsets

	A	B	C	D	E
1	word	1	2	best	comment
2	médico (geneeskundig, medisch)	El médico recetó antibióticos para tratar la infección de garganta de la paciente .	Se amplía el alquiler de Sextas para paliar las esperas en días de máxima afluencia de visitantes y , también en Sextas , se abre un nuevo centro médico .		
3	sentado (zittend, gezeten)	Durante la clase , los estudiantes estaban sentados en silencio mientras la profesora explicaba la lección .	Al entrar , varias personas están sentadas en una salita que hace las veces de recepción y pequeña librería improvisada .		
4	breve (beknopt, kort)	Antes de subir a la habitación , el viajero es invitado a una breve visita guiada por las dependencias de uso común .	La reunión fue muy breve , apenas duró quince minutos antes de que todos se retiraran .		
5	particular (bijzonder, speciaal)	La vivienda particular (y estudio) es una de las obras maestras del arquitecto y está abierta al público como museo .	Cada estudiante tiene su estilo particular de aprendizaje que debemos tener en cuenta en el aula .		
6	fundamental (essentieel, fundamenteel)	Los barcos que cruzan el Tajo de una orilla a otra tienen una importancia fundamental para la ciudad .	La participación activa de los estudiantes es fundamental para el éxito en el aprendizaje de un nuevo idioma .		
7	desconocido (onbekend)	El joven explorador se adentró en un territorio desconocido en busca de nuevas experiencias y conocimientos .	Por el momento se podrá ver parte de la colección , obras que forman dos exposiciones y que abren una puerta a mundos desconocidos de Newton .		
8	fuerte (hard, krachtig, luid, machtig, sterk)	Las dinámicas eran tan fuertes que , a veces , tenías la sensación de estar colocado .	La tormenta fue tan fuerte que provocó inundaciones en varias zonas de la ciudad .		
9	lento (langzaam, traag)	El ritmo de la vida en los pueblos es lento y al margen completamente del turismo masivo y de cualquier contaminación de la vida moderna .	El tren lento llegó a la estación con diez minutos de retraso , causando preocupación entre los pasajeros impacientes .		
10	intelectual (intellectueel)	Le gusta participar en discusiones intelectuales sobre filosofía y arte contemporáneo en su tiempo libre .	El restaurante Cacho , abierto hace un año y medio , se ha convertido en el favorito de modernos e intelectuales (Capitán Vigueras , 21 ; 954 53 03 95 .		
11	práctico (handig, praktisch)	La práctica totalidad de la vida en el perímetro interior se ha desplazado a los barrios residenciales del sur y a las ciudades satélite integradas como Overland Park .	El bolso grande y práctico que compré me ha sido muy útil para llevar todas mis cosas al trabajo .		

	A	B	C	D	E	F	G	H	I
1	word	1	2	3	4	5	best	worst	comment
2	político (aangetrouwd, politiek)	El discurso político del presidente generó opiniones divididas entre la población .	El discurso político del presidente fue muy persuasivo y tuvo gran impacto en la opinión pública .	Los de los partidos políticos acompañan a sus votantes en la otra vida .	La situación política en América Latina es muy complicada debido a diversos factores económicos y sociales .	Es importante estar informado sobre la situación político-social de un país para comprender su realidad y desarrollo .			
3	preparado (bereid, gereed)	El chef preparado sorprendió a todos con su exquisito menú degustación en el evento gastronómico de la ciudad .	Ana estaba muy preparada para la entrevista de trabajo , investigó sobre la empresa y practicó posibles respuestas a las preguntas .	En todos los casos sorprende la rapidez del servicio , teniendo en cuenta que sus cocinas no recurren a platos preparados (V gama) .	Aunque no lo crean , el restaurante preparado para la inauguración era un caos total , ¡ nada estaba en su lugar !	Los estudiantes deben estar preparados para el examen de gramática que tendremos la próxima semana en clase .			
4	útil (bruikbaar, dienstig, nuttig)	El paisaje que se extiende a los pies de la villa es útil y fascinante a un tiempo .	Es importante llevar herramientas útiles cuando se hace senderismo en la montaña para estar preparado ante cualquier imprevisto .	Es importante tener herramientas útiles para reparar cosas en casa y solucionar problemas cotidianos .	Es importante llevar contigo una linterna cuando te vayas de excursión porque es muy útil en caso de emergencia .	El diccionario es una herramienta útil para mejorar tu vocabulario en español .			
5	terrible (afschuwelijk, verschrikkelijk, vreselijk)	Anoche tuve un sueño terrible que me dejó asustado durante toda la mañana .	La película que vimos anoche fue realmente terrible , no pudimos ni terminar de verla .	Ayer tuvimos un examen de gramática y fue terrible , espero que en el próximo tengamos mejor suerte .	El terrible incendio destruyó por completo la casa familiar en cuestión de minutos .	El final de la guerra fue terrible para Dresde .			
6	convencido (overtuigd)	Por el contrario , los cubanos están convencidos de que todo el mundo les entiende .	Juan está convencido de que va a aprobar el examen porque ha estudiado mucho esta semana .	A pesar de las dificultades , él está convencido de que logrará sus objetivos en el curso de español .	Estoy convencido de que con esfuerzo y práctica , todos ustedes llegarán a ser hablantes de español muy competentes .	Ana está convencida de que su esfuerzo valdrá la pena y alcanzará sus metas a largo plazo en su carrera .			

CORPUS-BASED VERSUS Gen-AI

- Evaluation criteria students
 - Sentence is more definition than usage example
 - Sentence is context-dependent
 - Sentence is too difficult
- If all criteria met → base evaluation on needs as L2 learner
- Pairwise: indicate “best” (subset 1 to 4)
- Best-worst scaling: indicate “best” and “worst” (subset 5)

CORPUS-BASED VERSUS Gen-AI

- Formality check
 - Target item occurs in sentence (no “hallucination”)
 - Target item occurs in correct part of speech
 - Sentence is complete

RESULTS

	A	B	C	D	E
1	word	1	2	best	comment
2	médico (geneeskundig, medisch)	El médico recetó antibióticos para tratar la infección de garganta de la paciente .	Se amplía el alquiler de Sextas para paliar las esperas en días de máxima afluencia de visitantes y , también en Sextas , se abre un nuevo centro médico .		
3	sentado (zittend, gezeten)	Durante la clase , los estudiantes estaban sentados en silencio mientras la profesora explicaba la lección .	Al entrar , varias personas están sentadas en una salita que hace las veces de recepción y pequeña librería improvisada .		
4	breve (beknopt, kort)	Antes de subir a la habitación , el viajero es invitado a una breve visita guiada por las dependencias de uso común .	La reunión fue muy breve , apenas duró quince minutos antes de que todos se retiraran .		
5	particular (bijzonder, speciaal)	La vivienda particular (y estudio) es una de las obras maestras del arquitecto y está abierta al público como museo .	Cada estudiante tiene su estilo particular de aprendizaje que debemos tener en cuenta en el aula .		
6	fundamental (essentieel, fundamenteel)	Los barcos que cruzan el Tajo de una orilla a otra tienen una importancia fundamental para la ciudad .	La participación activa de los estudiantes es fundamental para el éxito en el aprendizaje de un nuevo idioma .		
7	desconocido (onbekend)	El joven explorador se adentró en un territorio desconocido en busca de nuevas experiencias y conocimientos .	Por el momento se podrá ver parte de la colección , obras que forman dos exposiciones y que abren una puerta a mundos desconocidos de Newton .		
8	fuerte (hard, krachtig, luid, machtig, sterk)	Las dinámicas eran tan fuertes que , a veces , tenías la sensación de estar colocado .	La tormenta fue tan fuerte que provocó inundaciones en varias zonas de la ciudad .		
9	lento (langzaam, traag)	El ritmo de la vida en los pueblos es lento y al margen completamente del turismo masivo y de cualquier contaminación de la vida moderna .	El tren lento llegó a la estación con diez minutos de retraso , causando preocupación entre los pasajeros impacientes .		
10	intelectual (intellectueel)	Le gusta participar en discusiones intelectuales sobre filosofía y arte contemporáneo en su tiempo libre .	El restaurante Cacho , abierto hace un año y medio , se ha convertido en el favorito de modernos e intelectuales (Capitán Vigueras , 21 ; 954 53 03 95 .		
11	práctico (handig, praktisch)	La práctica totalidad de la vida en el perímetro interior se ha desplazado a los barrios residenciales del sur y a las ciudades satélite integradas como Overland Park .	El bolso grande y práctico que compré me ha sido muy útil para llevar todas mis cosas al trabajo .		

PAIRWISE COMPARISONS

– Beginner group (full agreement)

Subset	Corpus-based = best	GenAI-based = best
1	1	38
2	1	39
3	0	36
4	2	35
Total	4 / 200	148 / 200

– Advanced group (full agreement)

Subset	Corpus-based = best	GenAI-based = best
1	1	28
2	1	28
3	0	33
4	1	28
Total	6 / 200	117 / 200

PAIRWISE COMPARISONS

– Beginner group (corpus-based, / 50)

Subset	No target item	Wrong POS	Incomplete	Definition	Context-dependent	Too difficult
1	0	1	2	0	16	11
2	0	1	2	0	13	12
3	0	0	0	0	15	11
4	0	0	1	0	14	9

– Beginner group (GenAI-based, / 50)

Subset	No target item	Wrong POS	Incomplete	Definition	Context-dependent	Too difficult
1	0	1	0	0	0	1
2	1	1	0	0	0	1
3	0	0	0	0	0	0
4	0	1	0	0	0	0

PAIRWISE COMPARISONS

– Advanced group (corpus-based, / 50)

Subset	No target item	Wrong POS	Incomplete	Definition	Context-dependent	Too difficult
1	0	3	0	0	14	23
2	0	1	0	1	16	21
3	0	1	0	0	15	19
4	0	0	0	0	14	23

– Advanced group (GenAI-based, / 50)

Subset	No target item	Wrong POS	Incomplete	Definition	Context-dependent	Too difficult
1	0	2	0	3	1	3
2	0	2	0	1	0	0
3	0	3	0	5	1	0
4	0	0	0	3	1	2

	A	B	C	D	E	F	G	H	I
1	word	1	2	3	4	5	best	worst	comment
2	político (aangetrouwd, politiek)	El discurso político del presidente generó opiniones divididas entre la población .	El discurso político del presidente fue muy persuasivo y tuvo gran impacto en la opinión pública .	Los de los partidos políticos acompañan a sus votantes en la otra vida .	La situación política en América Latina es muy complicada debido a diversos factores económicos y sociales .	Es importante estar informado sobre la situación político-social de un país para comprender su realidad y desarrollo .			
3	preparado (bereid, gereed)	El chef preparado sorprendió a todos con su exquisito menú degustación en el evento gastronómico de la ciudad .	Ana estaba muy preparada para la entrevista de trabajo , investigó sobre la empresa y practicó posibles respuestas a las preguntas .	En todos los casos sorprende la rapidez del servicio , teniendo en cuenta que sus cocinas no recurren a platos preparados (V gama) .	Aunque no lo crean , el restaurante preparado para la inauguración era un caos total , ¡ nada estaba en su lugar !	Los estudiantes deben estar preparados para el examen de gramática que tendremos la próxima semana en clase .			
4	útil (bruikbaar, dienstig, nuttig)	El paisaje que se extiende a los pies de la villa es útil y fascinante a un tiempo .	Es importante llevar herramientas útiles cuando se hace senderismo en la montaña para estar preparado ante cualquier imprevisto .	Es importante tener herramientas útiles para reparar cosas en casa y solucionar problemas cotidianos .	Es importante llevar contigo una linterna cuando te vayas de excursión porque es muy útil en caso de emergencia .	El diccionario es una herramienta útil para mejorar tu vocabulario en español .			
5	terrible (afschuwelijk, verschrikkelijk, vreselijk)	Anoche tuve un sueño terrible que me dejó asustado durante toda la mañana .	La película que vimos anoche fue realmente terrible , no pudimos ni terminar de verla .	Ayer tuvimos un examen de gramática y fue terrible , espero que en el próximo tengamos mejor suerte .	El terrible incendio destruyó por completo la casa familiar en cuestión de minutos .	El final de la guerra fue terrible para Dresde .			
6	convencido (overtuigd)	Por el contrario , los cubanos están convencidos de que todo el mundo les entiende .	Juan está convencido de que va a aprobar el examen porque ha estudiado mucho esta semana .	A pesar de las dificultades , él está convencido de que logrará sus objetivos en el curso de español .	Estoy convencido de que con esfuerzo y práctica , todos ustedes llegarán a ser hablantes de español muy competentes .	Ana está convencida de que su esfuerzo valdrá la pena y alcanzará sus metas a largo plazo en su carrera .			

BEST-WORST SCALING

- Beginner group (full agreement, / 50)

Subset	Corpus-based	General	General + target aud	General + target aud + criteria	Few shot
5_{best}	0	0	6	3	4
5_{worst}	28	0	0	0	2

- Beginner group (≥ 1 agreement, / 50)

Subset	Corpus-based	General	General + target aud	General + target aud + criteria	Few shot
5_{best}	3	23	24	24	24
5_{worst}	47	5	6	6	10

BEST-WORST SCALING

- Advanced group (full agreement, / 50)

Subset	Corpus-based	General	General + target aud	General + target aud + criteria	Few shot
5_{best}	1	2	0	3	2
5_{worst}	26	0	1	0	1

- Advanced group (≥ 1 agreement, / 50)

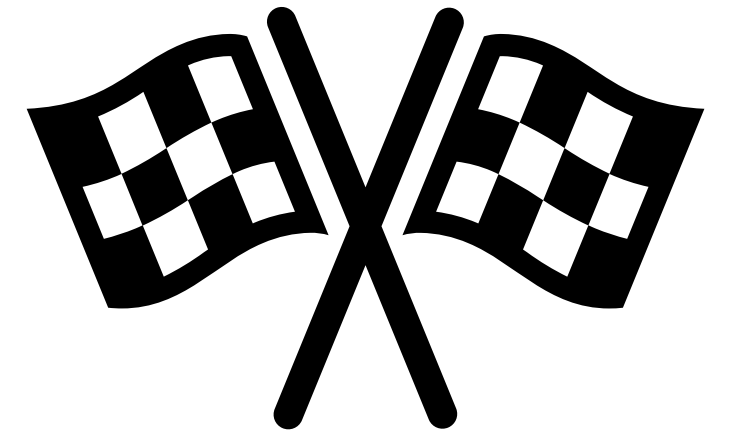
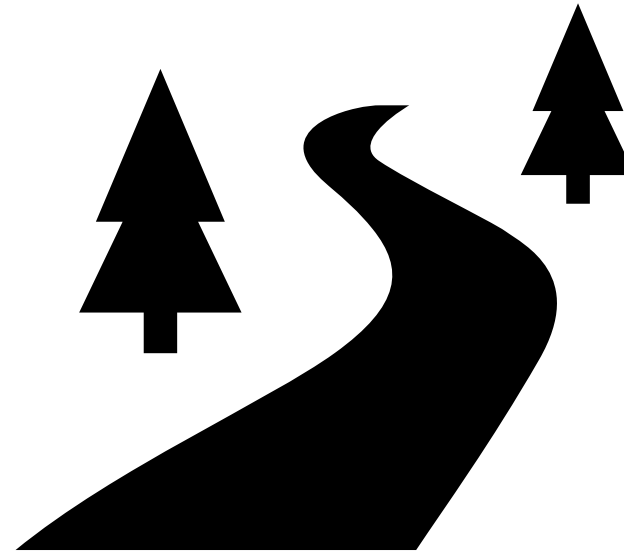
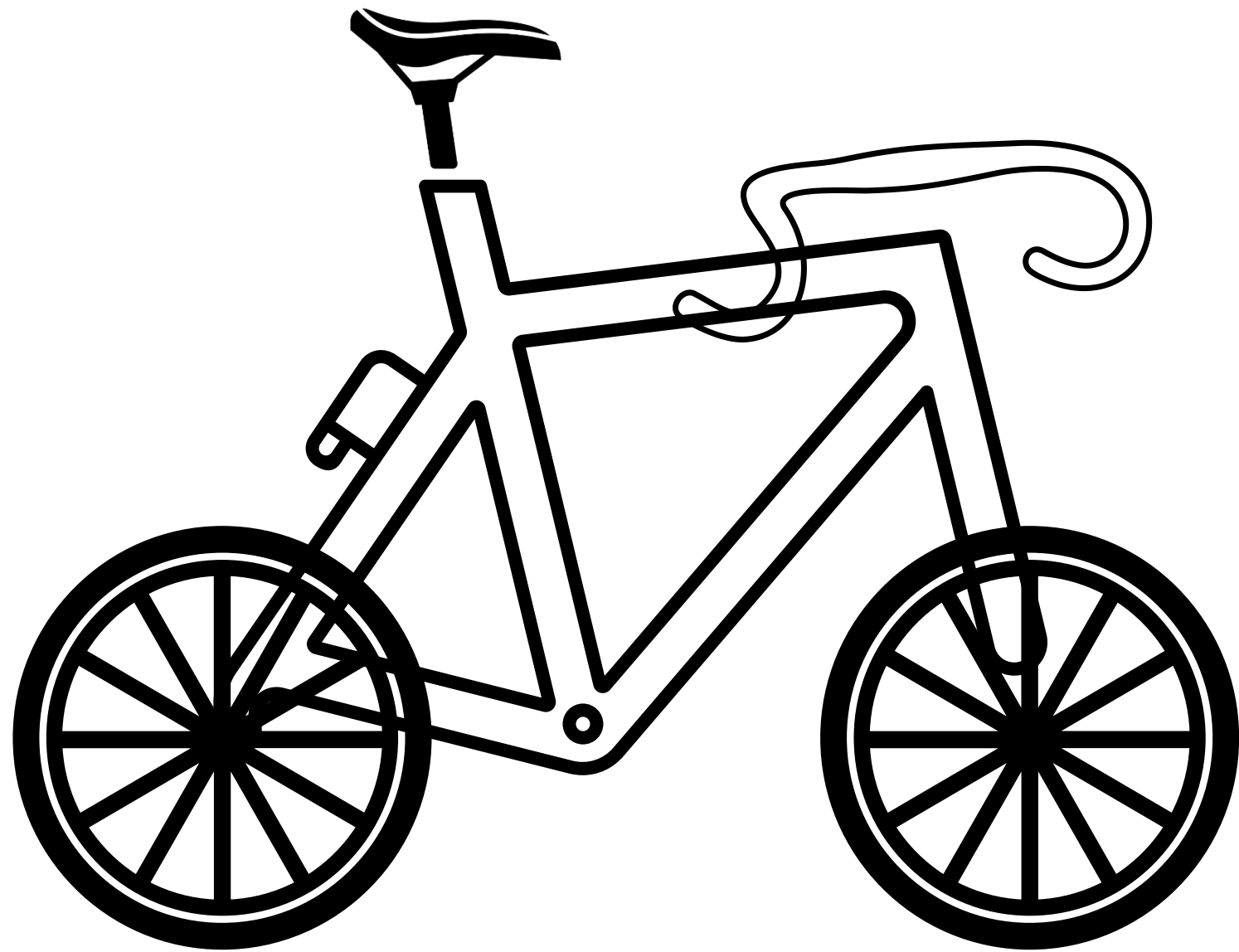
Subset	Corpus-based	General	General + target aud	General + target aud + criteria	Few shot
5_{best}	4	32	26	31	28
5_{worst}	45	6	9	2	8

CONCLUSION

CONCLUSION

- Novel framework to select **pedagogically suitable** sentences from corpora for L2 Spanish
- **GenAI more suitable** source of example sentences than corpus-based examples
- **BUT:** not always
 - “**hybrid**” approach as combination of best of both worlds?

LET'S GO FOR A RIDE



REFERENCES

- Cook, G. (2001). ‘The philosopher pulled the lower jaw of the hen’. Ludicrous invented sentences in language teaching. *Applied Linguistics*, 22(3), 366–387.
<https://doi.org/10.1093/applin/22.3.366>
- Davies, M., & Hayward Davies, K. (2018). *A frequency dictionary of Spanish: Core vocabulary for learners* (2nd ed.). Routledge.
- Frankenberg-Garcia, A. (2012). Learners’ Use of Corpus Examples. *International Journal of Lexicography*, 25(3), 273–296. <https://doi.org/10.1093/ijl/ecs011>
- Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., & Rychlý, P. (2008). GDEX: Automatically finding good dictionary examples in a corpus. *Proceedings of the XIII EURALEX International Congress*, 1, 425–432.
- Laufer, B., & Shmueli, K. (1997). Memorizing New Words: Does Teaching Have Anything To Do With It? *RELC Journal*, 28(1), 89–108.
<https://doi.org/10.1177/003368829702800106>
- Pilán, I., Volodina, E., & Borin, L. (2016). Candidate sentence selection for language learning exercises: From a comprehensive framework to an empirical evaluation. *Revue Traitement Automatique Des Langues*, 57(3), 67–91.

Jasper Degraeuwe

PhD researcher

DEPARTMENT OF TRANSLATION,
INTERPRETING AND COMMUNICATION

E Jasper.Degraeuwe@UGent.be

www.ugent.be

 Universiteit Gent

 @ugent

 @ugent

 Ghent University

PROMPT TYPES (1)

Prompt ID	Prompt text
SYS1	You are a teacher of Spanish as a foreign language.
SYS2	You are a teacher of Spanish as a foreign language to a beginner/lower-intermediate group of university students who have been studying Spanish for one year.
SYS3	You are a teacher of Spanish as a foreign language to an upper-intermediate/advanced group of university students who have been studying Spanish for three years.
USR1	Write a sentence between 15 and 30 words in Spanish that presents an authentic usage of the Spanish [POS] '[WORD]', a vocabulary item that has to be learnt by your students. The sentence should not be a definition of the word.
USR2	Write a sentence between 15 and 30 words in Spanish that presents an authentic usage of the Spanish [POS] '[WORD]', a vocabulary item that has to be learnt by your students. The sentence should not be a definition of the word. The sentence should be well-formed and context-independent, it should be tailored to the proficiency level of your students, and it should contain phrases that frequently co-occur with the target item '[WORD]'.
USR3	Write a sentence between 10 and 30 words in Spanish that presents an authentic usage of a Spanish vocabulary item that has to be learnt by your students: word=diseño; part of speech=noun; sentence=Para hacer un buen diseño de un mueble hay que pensar en su utilidad. ### word=comprometer; part of speech=verb; sentence=Sus revelaciones comprometían en el caso de corrupción a otras dos organizaciones. ### word=dramático; part of speech=adjective; sentence=Toda la prensa se hace eco del dramático caso de la niña desaparecida. ### word=[WORD]; part of speech=[POS]; sentence=

PROMPT TYPES (2)

Prompt type	System role	User role	Subset
Beginner			
ZS-GEN	SYS1	USR1	1
ZS-GEN+AUD	SYS2	USR1	2
ZS-GEN+AUD+CRIT	SYS2	USR2	3
FEWSHOT	SYS2	USR3	4
Advanced			
ZS-GEN	SYS1	USR1	6
ZS-GEN+AUD	SYS3	USR1	7
ZS-GEN+AUD+CRIT	SYS3	USR2	8
FEWSHOT	SYS3	USR3	9

Subset	IAA (α)	ZS-G	ZS-G+A	ZS-G+A+C	FEWSH	ALL
1	0.7	✓				
2	0.72		✓			
3	0.66			✓		
4	0.62				✓	
5 _{best}	0.29					✓
5 _{worst}	0.61					✓
Avg	0.6	✓	✓	✓	✓	✓
6	0.6	✓				
7	0.58		✓			
8	0.55			✓		
9	0.65				✓	
10 _{best}	0.22					✓
10 _{worst}	0.71					✓
Avg	0.55	✓	✓	✓	✓	✓