

# Le rôle du corpus dans l'apprentissage des collocations : corpus d'apprenants et corpus pour apprendre

**M. Alonso Ramos**



UNIVERSIDADE DA CORUÑA

*Séminaire DéLiCorTal*  
Université Grenoble-Alpes  
14 de février 2020

# Introduction

## ● Object

Collocations (au sens de Hausmann, Mel'čuk):

*dar un **paseo**/ faire une **promenade**/ take a **walk**  
**fumador empedernido** / gros **fumeur**/ heavy **smoker***

## ● Motivation

consensus sur des collocations:

**difficiles** pour les apprenants de L2 (**anglais**)

**MAIS**

et avec avec les apprenants d' **espagnol**?

## ● Objectif

- quels besoins collocationnels des apprenants
- les ressources d'apprentissage satisfont ces besoins?

# Introduction

- **Comment obtenir cette information**

- corpus d'apprenants pour le diagnostic

- Corpus de aprendices de español, CEDEL2**

- fragment annoté avec des collocations

- nouvelle ressource: apprendre avec un corpus

- HARenEs**

- **Objectifs de cet exposé**

- 1. survol des résultats après avoir annoté CEDEL2

- 2. évaluer l'effectivité de HARenEs comme une ressource d'apprentissage de collocations

# Plan de l'exposé

- 1. Notion de collocation dans notre cadre théorique**
- 2. Collocations dans un corpus d'apprenants**
- 3. Assistant de collocations en ligne: HARenEs**
- 4. Usage de HARenEs par les apprenants**
- 5. Conclusions**

# 1. Notion de Collocation

# Notion de collocation

- **Notion clé**

cooccurrence lexicale restreinte

- **Lexicologie Explicative et Combinatoire (Mel'čuk et al.):**

Unité phraséologique  $L_1L_2$

$L_1$  = **base** de la collocation

$L_2$  = **collocatif**

- **Collocation: un syntagme compositionnel**

' $L_1 L_2$ ' = ' $L_1$ ' + ' $L_2$ '

**mais pas libre**

- la **base** est sélectionnée après **son sens**

- le **collocatif** est sélectionné sous le **contrôle de la base**

# Choix lexicaux

## ✓ **CONTRÔLÉS sémantiquement**

la base est sélectionnée pour exprimer un sens indépendamment d'autres UL

'síntoma de enfermedad que consiste en la elevación de la temperatura del cuerpo (...)' ==> **FIEBRE**

## ✓ **CONTRÔLÉS lexicalement**

le collocatif es sélectionné pour exprimer un sens donné, mais sous le contrôle d'une autre UL déjà choisie

'intenso' dit de **FIEBRE**==>**ALTA**

## **2. Collocations dans un corpus d'apprenants d'espagnol**



## 2. Collocations dans CEDEL2

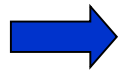
● Pour offrir de bon outils d'apprentissage, il faut connaître la production collocationnelle des apprenants

● Un coup d'oeil à CEDEL2

<http://cedel2.learnercorpora.com/>

montre que les erreurs collocationnelles sont très différentes

- *salvar* dinero 'sauver l'argent' (au lieu de ahorrar dinero)
- recibir un *llamo* 'recevoir un appel' (au lieu de recibir una llamada)
- *asistir* la universidad, 'assister l'université' (au lieu de asistir a la universidad)



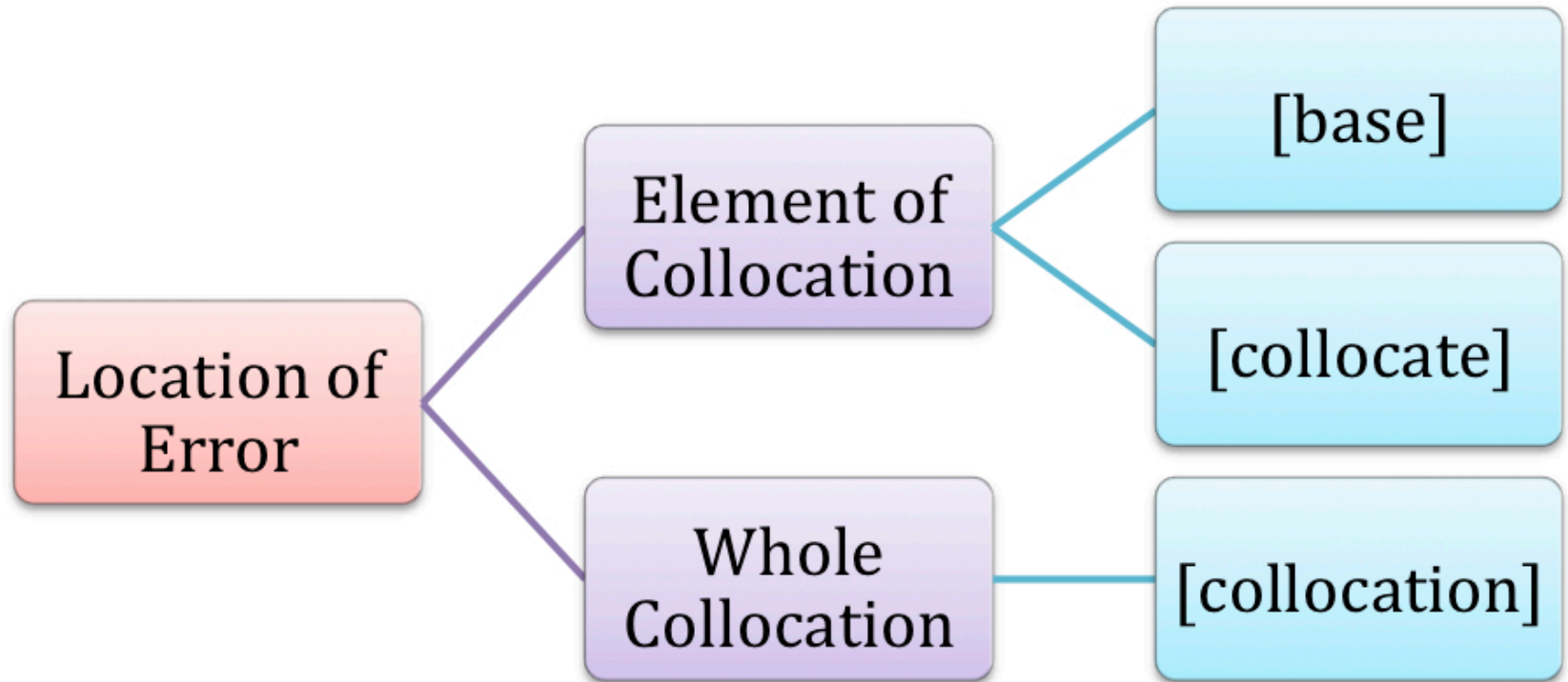
● Une classification plus détaillée des erreurs collocationnelles est nécessaire

## 2. Collocations dans CEDEL2

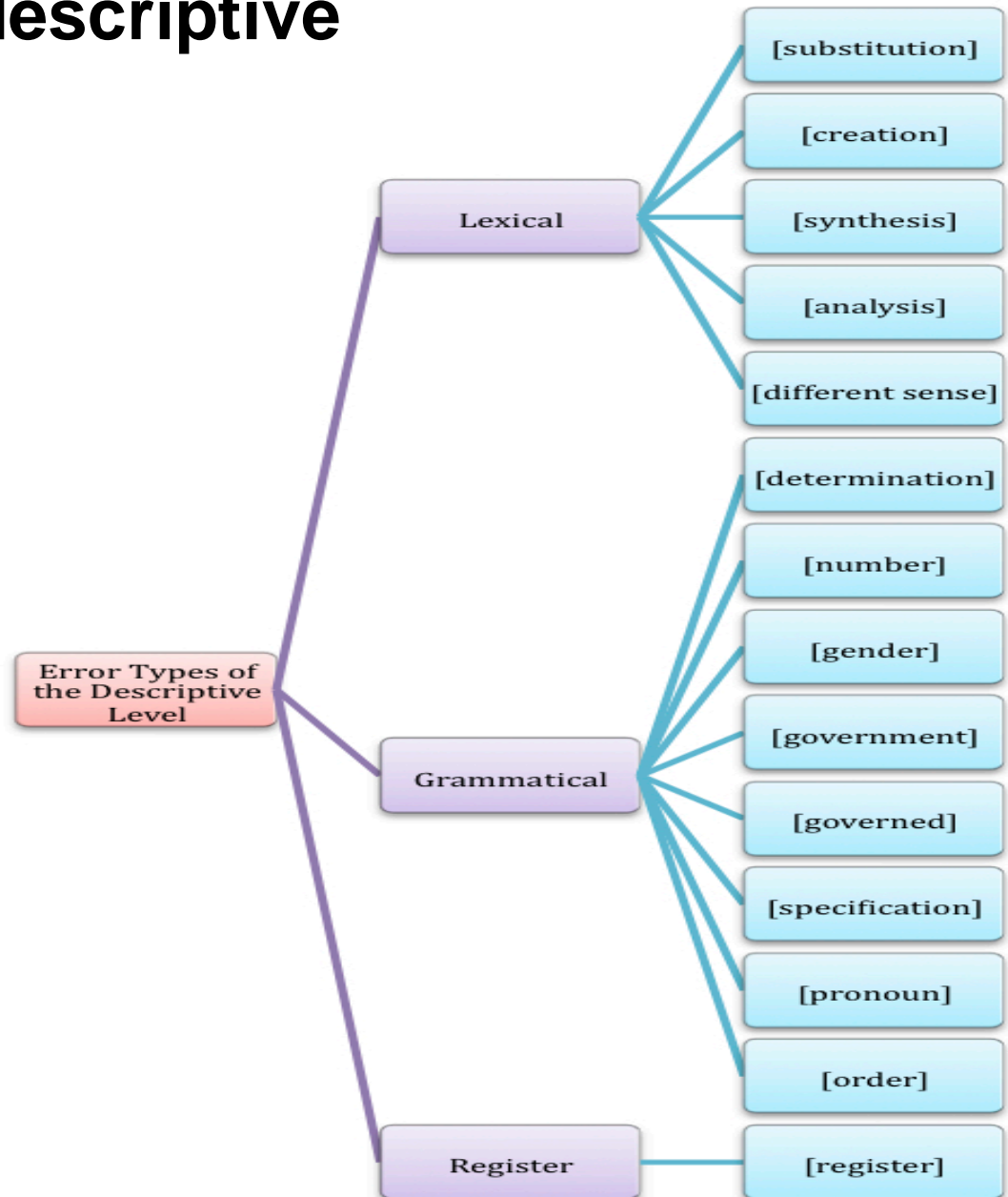
Typologie tridimensionnelle des erreurs collocationnelles

- localisation
- descriptive
- explicative

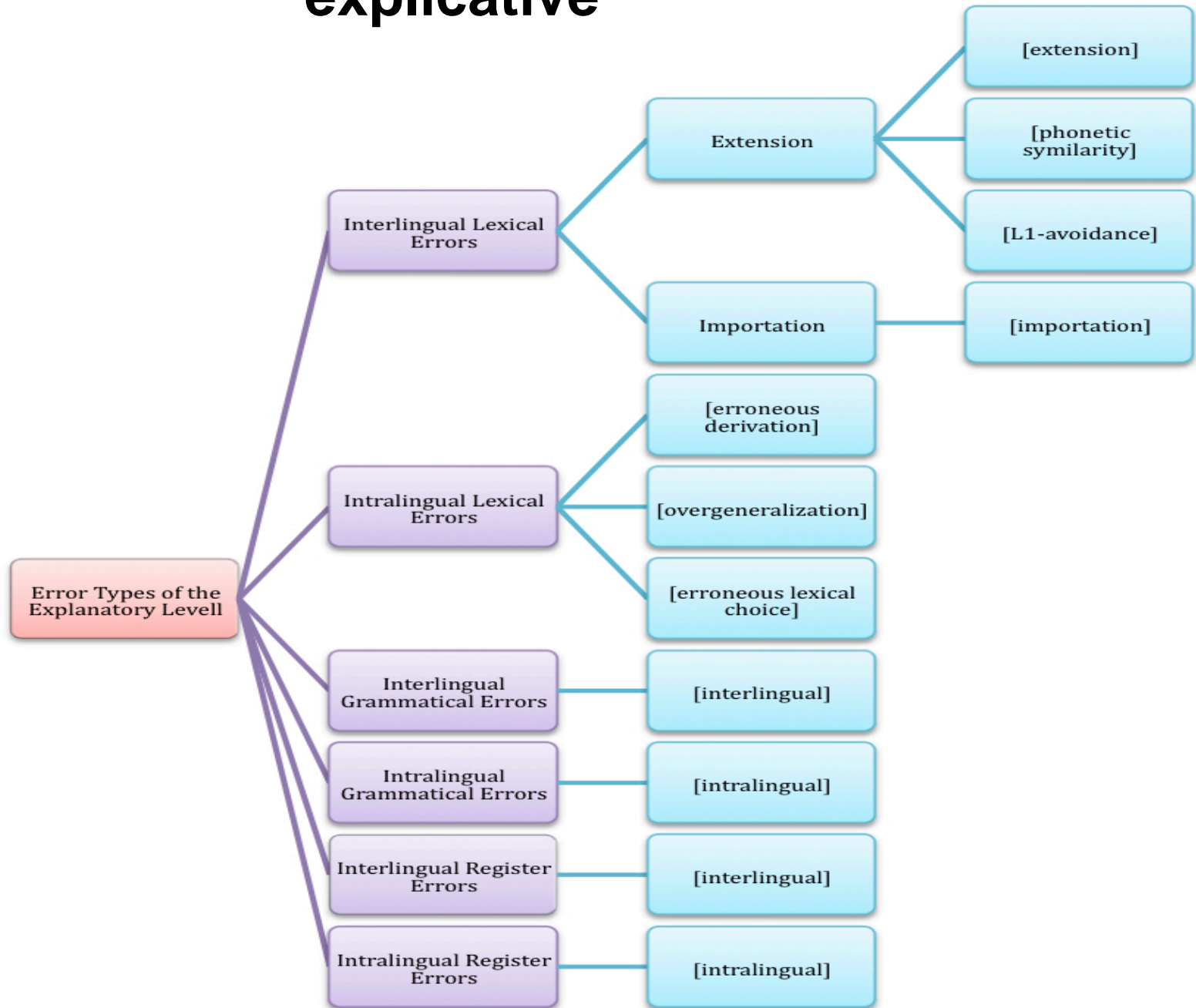
## Dimension: localisation



# Typologie d'erreurs: dimension descriptive



# Typologie d'erreurs: dimension explicative



# Illustration d'erreurs lexicales interlingues (affectant la base ou le collocatif)

		EXPLANATORY CRITERION			
		Interlingual Errors			
LOCATION	DESCRIPTIVE CRITERION	Importation	Extension		
				Phonetic similarity	L1 avoidance
BASE	Substitution	hicimos <u>wakeboarding</u>	<u>juego</u> de fútbol (from <i>game</i> )	Doy una <u>marca</u> (instead of <i>poner una nota</i> )	<u>Estado</u> económico (from <i>economic situation</i> )
	Creation	recibí un <u>llamo</u> (analogy with <i>call</i> )			
COLLOCATE	Substitution		<u>gastar</u> todo el año (from <i>spend</i> ) <u>tomé</u> examen (from <i>take</i> ) <u>Doy</u> una marca (from <i>give</i> )	lengua <u>maternal</u> (from <i>maternal language</i> ) <u>capturar</u> la atención (from <i>capture</i> )	<u>cambiar</u> a la verdadera religión (Eng. <i>convert to a religión</i> ) <u>acudir</u> el teléfono (Eng. <i>attend the telephone</i> )
	Creation				

# Illustration d'erreurs *lexicales intralingues* (affectant la *base* ou le *collocatif*)

EXPLANATORY CRITERION			
Intralingual Errors			
LOCATION	Erroneous derivation	Overgeneralization	Err. lexical choice
BASE	recibí un <u>llamo</u> (analogy with <i>paseo, canto, salto..</i> )		Me da el <u>sentido</u> (instead of <i>la sensación</i> )
COLLOCATE	enseñanza <u>segundaria</u> (from <i>segundo</i> )	<u>hacer</u> citas (instead of <i>concertar</i> )  <u>malos</u> efectos (instead of <i>nocivos, dañinos</i> )	<u>toman</u> puestos (instead of <i>ocupar puestos</i> )  <u>escribir</u> el examen (instead of <i>hacer</i> )

# Illustration d'erreurs **grammaticales** (affectant la **base** ou le **collocatif**)

		EXPLANATORY CRITERION	
LOCATION	DESCRIPTIVE CRITERION	Interlingual errors	Intralingual errors
BASE	determination	<ul style="list-style-type: none"> <li>• tienen <u>el</u> derecho de (from <i>to have the right to</i>)</li> <li>• terminé escuela (from <i>to finish school</i>)</li> </ul>	<ul style="list-style-type: none"> <li>• tomé examen (instead of <i>tomé un examen</i>)</li> </ul>
	number		<ul style="list-style-type: none"> <li>• hicimos los <u>esquí</u>s del agua (instead of <i>esquí de agua</i>)</li> <li>• tienen <u>prejuicio</u> (instead of <i>prejuicios</i>)</li> </ul>
	gender		<ul style="list-style-type: none"> <li>• <u>días</u> festivos (instead of <i>días festivos</i>)</li> </ul>
	government	<ul style="list-style-type: none"> <li>• tengo interés <u>in</u></li> <li>• tengo razones <u>por</u></li> </ul>	
	specification		<ul style="list-style-type: none"> <li>• probar comida (an article or an adjective should be used to specify the base, e.g. <i>probar comida exótica</i>)</li> </ul>
COLLOCATE	government	<ul style="list-style-type: none"> <li>• hablando <u>al</u> teléfono (from Italian, the student's L2: <i>parlare al telefono</i>)</li> <li>• asisto un juego de fútbol (from <i>assist a game of...</i>)</li> <li>• entró la universidad (from <i>enter university</i>)</li> </ul>	<ul style="list-style-type: none"> <li>• montó el autobús (instead of <i>montar en..</i>)</li> </ul>
	pronoun		<ul style="list-style-type: none"> <li>• muero de ganas (instead of <i>por me muero de...</i>)</li> <li>• la película se trata.. (instead of <i>la película trata</i>)</li> </ul>



# Annotation des collocations avec knowtator

Colocate\_demo1 - Protégé 3.3.1 (file:\C:\Documents%20and%20Settings\Usuario\Escritorio\proyecto\_demo\Colocate\_demo1.ppr), Protégé Files (.pont and .pins)

File Edit Project Window Knowtator Help

Text source collection: coleccion nueva Filter: Ver colocaciones

Knowtator Classes

annotation schema

- Colocación Correcta (7)
- Colocación Incorrecta (14)
- Quasilocución
- Base
- Colocativo

text source: B.txt filter: Ver colocaciones

El año pasado durante las vacaciones del verano, trabajaba para los directores de unos grupos musicales de música punk. El trabajo era en Los Angeles, donde **estaba la universidad**. Empecé mi experiencia con la compañía, se llama Villam Artist Management, como voluntaria, porque **tenía ganas de** aprender eso tipo de trabajo. La compañía compartía un edificio en Hollywood con otra compañía que **produce discos** porque uno de nuestros grupos que **tiene el más éxito ha producido dos discos** **debajo de la marca** de la otra compañía (Side One Dummy Records). La compañía para que trabajara es muy pequeña. Ellos consisten de dos empleados- uno es él con el más poder, el **gran jefe** bromeábamos, y la otra es su única empleada que él no le ha despedido. El **problema es que él pide** demasiado de sus empleadas. El es un **gran hombre de negocios** y **comanda respeto de** todos que le conocen, pero sus **estándares** son demasiado **altos**. Él quiere que sus empleadas **sepan un poco** de trabajar tan **fuerte** como él, y también, no acepta "no" para una respuesta a lo que pide. Cuando llegué a la compañía, había una secretaria sustituta porque él la ha despedido a su secretaria de dos años porque ella pedía una jornada más corta y él pensaba que esa pedida significaba que ella era perezosa. El próximo secretario era el primo de la otra empleada. Eso nuevo trabajador descasaba pronto de la cantidad de trabajo muy rápido y de los **estándares** **altos** que **tenía** Gary (el jefe). **Por la terminación de** **servicio**, yo trabajaba como secretario. El me pagaba para esa trabaja, aunque era muy difícil **mantener el ritmo** que él me ha prometido. En mi nueva posición, yo **me planea** de viajar para los grupos, **acudí al teléfono** e **hice citas** para conferencias con otras compañías para Gary. También seguí con mis responsabilidades de antes- que todavía tenía que buscar nuevas maneras para hacer más eficiente la compañía. El trabajo era muy difícil—de hecho, todo de ese verano era muy difícil. Llegué a la oficina por la mañana a las 8 y salí a las 5, para que pudiera ir a mi otro trabajo como camarera en un restaurante. Por lo general, terminé **la medianoche** y como soy joven, y nada es más importante en este punto de mi vida que mis amigos y mi vida social, salí con ellos hasta muy temprano... yo sé que la falta de sueño es culpa mía, pero sin embargo, era difícil. Lo bueno del trabajo para Villam Artists es que Gary no me despidió tampoco quería ir yo cuando le dejé. Era un **gran chico** cuando Gary me invitó a almorzar conmigo y me contó que siempre habrá sitio para mi en su compañía si quería regresar para él después de escuela.

DELETE **comanda respeto de**

span edit: << >> <<< >>>

annotated class

- Colocación Incorrecta

slots of annotated class (14 values)

- 12-función Léxica
  - Caus
  - Func
- 1
- 2-Error1: Localización Error
  - colocativo
- 21-Error1: Tipo Error Descriptivo
  - Lex\_miembro\_colocacion
- 22-Error1: Tipo Error Explicativo

Symbol selection

- importacion
- importacion
- extension
- ext semejanza\_fonetica
- ext evasion\_1\_1
- derivacion\_erronea
- comodin
- falsa\_eleccion
- gram\_transferencia

annotation set membership

comanda respeto de

# Quelques chiffres

- 1) Du total des collocations annotées, 76% sont correctes et 24% incorrectes
  - 2) 57% des erreurs sont lexicales, 43% grammaticales
  - 3) 52% affectent le collocatif, 35% la base et 13% la collocation au complet
  - 4) Le 39% des erreurs lexicales sont dues à l'extension
  - 5) L'erreur grammaticale la plus fréquente concerne le régime
- **Bref**, les apprenants
    - emploient les collocations plutôt correctement
    - commettent différents types d'erreurs collocationnelles

# Pour plus d'information

**A corpus study of Spanish  
as a Foreign Language  
learners' collocation  
production**

[Orsolya Vincze](#)

[Marcos García-Salido](#)

[Ana Orol](#)

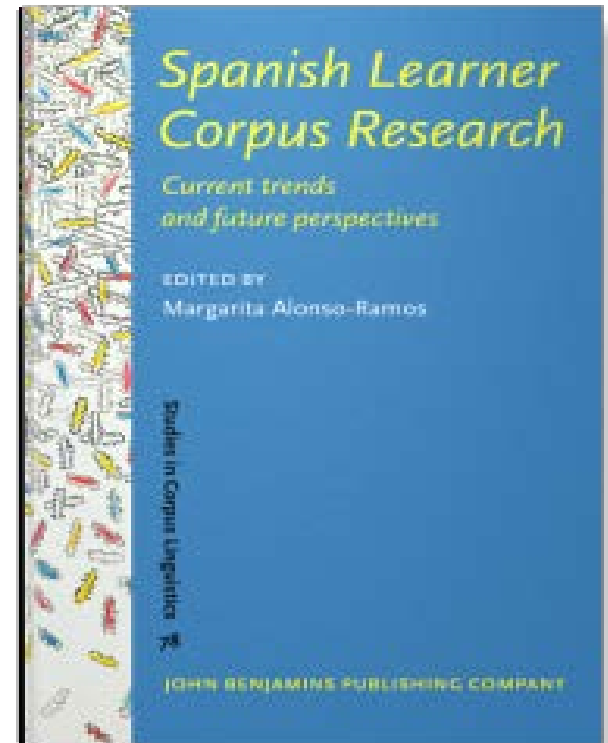
[Margarita Alonso-Ramos](#)

*Chapter published in:*

[Spanish Learner Corpus  
Research: Current trends  
and future perspectives](#)

Edited by Margarita Alonso-  
Ramos [[Studies in Corpus  
Linguistics](#) 78] 2016

► pp. 299–331



# Ressource idéale

## pour aider les apprenants

- Dictionnaire?
- Corpus?
- Un autre type d'outil?
  - corpus-driven lexicography → “lexically-driven corpora”
  - limites flous entre dictionnaire et corpus
  - a ***word sketch*** is “output from the corpus, making it corpus-like, but it is a condensed summary of what was found there, making it dictionary-like.” (Kilgarriff 2009)

## Nouvel outil pour apprendre des collocations

- *Herramienta de ayuda a la redacción en español* (HARenEs)

### **3. Assistant de collocations en ligne: HARenEs**

# HARenEs: componentes



Herramienta de ayuda a la redacción en español:  
Procesamiento de colocaciones



[Homepage](#) [Blog](#) [Articles](#) [Photos](#) [Information](#) [Friends](#) [About](#) [Contact](#)



## BUSCAR COLOCACIONES

Este apartado permite introducir una palabra para comprobar con qué otras palabras se combina para formar colocaciones. El resultado será una lista de colocaciones organizadas según su estructura sintáctica o su contenido semántico.



## COMPROBAR COLOCACIONES

En este apartado puedes comprobar si una colocación es correcta o incorrecta ilustrando el resultado mediante una lista de ejemplos. Además de consultar una colocación aislada, también puedes introducir un texto para que la interfaz identifique y compruebe las colocaciones que contiene.



## MI DICCIONARIO

Puedes registrarte para crear un diccionario personal de colocaciones. En este apartado se irá creando una lista constituida por las colocaciones que vayas incorporando bien manualmente, bien desde los apartados de **Comprobar** y **Buscar colocaciones**.



# HARenEs: vérifier des collocations

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

Universidade da Coruña x Zimbra: Bandeja de entrad... x HARenES Project: Check x +

harenes.taln.upf.edu/CakeHARenEs/check

**HARenES**  
*Aprende colocaciones!*

Herramienta de ayuda a la redacción en español:  
Procesamiento de colocaciones

taln

Inicio Comprobar colocaciones Buscar colocaciones Mi diccionario Acerca de

### Comprobar colocación

Introduce una colocación para comprobar si es correcta  
(p.ej. *dar un paseo* o *tomar un paseo*)

poner la culpa ✖

Comprobar

Check\_Collocation\_Incorrect

### Sugerencias

echar culpa	<a href="#">Ver ejemplos</a> <span>+</span> <span>✓</span> <span>✗</span>
tener culpa	<a href="#">Ver ejemplos</a> <span>+</span> <span>✓</span> <span>✗</span>
expiar culpa	<a href="#">Ver ejemplos</a> <span>+</span> <span>✓</span> <span>✗</span>
entonar culpa	<a href="#">Ver ejemplos</a> <span>+</span> <span>✓</span> <span>✗</span>
cargar culpa	<a href="#">Ver ejemplos</a> <span>+</span> <span>✓</span> <span>✗</span>
Show more..	

### Collocation Valid

This section validates a collocation inserted by the user.

If the collocation is correct, the system shows a tick sign ✓ and a plus sign + that allows the user to add the collocation to the dictionary.

If the collocation is incorrect, the system shows a cross sign ✗ (see further explanation in the suggestions section). The user can validate the suggestions. The validated suggestions are stored into the user dictionary. Also, the user can link the collocation inserted to a...

ES 13:48 05/05/2015

# Corrections suggérées par HARenES

## Incorrecte

acudir el teléfono

cambiar al cristianismo

comandar respeto

comer café

gastar un año

hacer un paso

escribir un examen

lograr gol

-

## Reconnue

+

+

+

+

+

+

+

+

## Proposée

atender el teléfono

convertirse al cristianismo

imponer respeto

tomar café

pasar un año

dar un paso

pasar un examen

meter gol



# HARenEs: requête du collocatif

The screenshot shows a web browser window with the HARenEs application. The address bar displays the URL `haren.es.tal.n.upf.edu/CakeHARenEs/search`. The search bar contains the text "Buscar".

On the left side, there is a search filter panel with the following options:

- Lemma:
- Categoria:
- Combinado con...:
- Mostrar grafo

Below the filter panel is a "Buscar" button.

The main content area is titled "Results" and displays a list of collocates:

- concertar cita [Ver ejemplos](#) (+) (✓) (✗)
- dar cita [Ver ejemplos](#) (+) (✓) (✗)
- cancelar cita [Ver ejemplos](#) (+) (✓) (✗)
- afrontar cita [Ver ejemplos](#) (+) (✓) (✗)

The right side of the page is a large, empty, light-colored area, possibly a placeholder for a graph or additional results.

The Windows taskbar at the bottom shows the system tray with the date and time: 19:46, 29/01/2016.

# HARenEs: dictionnaire personnel

*Mi diccionario - Ver entrada* ⓘ

## Echar siesta

[Editar](#) [Eliminar](#) [Añadir a lista](#)

**Base:** siesta (Nombre)

**Colocativo (lemma):** echar

**Colocativo (flexionado):** echar

**Patrón:** verbo + ~

**Regimen:**

**Glosa:**

**Nota:**

**Ejemplos**

[Añadir ejemplo](#)

[← Back](#)

## **4. Emploi de HARenEs par des apprenants**

# Étude expérimentale

## Objectif principal

Tester l'effectivité de HARenES (hybride corpus et dictionnaire) avec des apprenants d'espagnol

## Deux expériences

- Groupe d'apprenants **japonais** au Japon (G1)
  - 45 participants (niveau entre B1 et B2)
  - Sousgroupe 1.1: **sans** ressources et **avec** HARenES
  - Sousgroupe 1.2: **sans** ressources et **avec** leurs dicos
- Groupe d'apprenants **Erasmus et chinois** en Espagne (G2)
  - 84 participants (niveau entre B1 et B2)
  - **sans** ressources et **avec** HARenES

## Tous les détails

*International Journal of Lexicography*, 32(4):480-497, 2019. ISSN 0950-3846.  
DOI [10.1093/ijl/ecz016](https://doi.org/10.1093/ijl/ecz016), [Link for free access](#)

# Instrument

- Questionnaire du profil et du niveau linguistique de l'apprenant
- Questionnaire: 10 N+Adj et 10 V+N

## Completa con la palabra que falta

En esta ocasión responde con la ayuda de HARENES.

Con esta herramienta puedes comprobar si una combinación de palabras es adecuada haciendo clic en COMPROBAR COLOCACIONES (<https://goo.gl/SQ6CL5>) o encontrar un adjetivo o un verbo apropiado en BUSCAR COLOCACIONES (<https://goo.gl/0TMgLH>).

En COMPROBAR COLOCACIONES tienes que escribir las dos palabras, cada una de ellas en la forma que buscarías en un diccionario.

Ejemplo:

Lucía es una AMIGA ... . Tenemos mucha confianza.

Posible respuesta: ÍNTIMA

Comprobar colocación: AMIGO ÍNTIMO

En BUSCAR COLOCACIONES, tienes que escribir el nombre en "Lemma" y elegir "Adjetivo" o "Verbo" en "Combinar con".

La prisa hace que elijamos COMIDA ... antes que comida casera.

Tu respuesta

---

# Résultats des expériences

	Moyenne de bonnes réponses	
	G1	G2
Avec H	13.57	15.44
Sans H	6.24	7.54
Avec D	12.83	
Sans D	6.21	

- Au moins, HARenES est **si effectif** que les dicos conventionnels même sans entraînement
- Les deux groupes emploient **plus le mode Dico** que le mode Corpus
- Regardent peu les exemples: 63% dans le G2
- Pas suffisamment convivial

# Évaluation de résultats

- **Correct** dans HARenEs, **incorrect** dans le contexte:  
*Juan es demasiado ambicioso y no puede #cruzar todas las metas que se propone*  
'Juan est trop ambitieux et il ne peut pas #traverser les lignes d'arrivée qu'il se propose'

*meta:*

1. 'ligne d'arrivée' → cruzar 'traverser'
2. 'but, objectif' → alcanzar 'atteindre'

Consultation insuffisante des exemples peut-être à cause de surestimation de leur connaissance

- **Incorrect** dans HARenES, **correcte** dans le contexte:  
*lluvia ininterrumpida* 'pluie ininterrompue'  
*éxito increíble* 'succès incroyable'

Peut-être pas dans le corpus ou pas suffisamment relevant pour les mesures d'association

**Globalement**, HARenEs est utile pour les apprenants dans leur production collocationnelle

## **5. Conclusions**



## 5. Conclusions

À propos des Collocations dans un Corpus d'apprenants

- 1) Erreurs collocationnelles ne sont pas homogènes
- 2) Une typologie d'erreurs détaillée est utile pour fournir matériel d'enseignement qui supporte l'apprentissage actif
- 3) La typologie est indépendante des langues

## 5. Conclusions (2)

À propos du corpus comme ressource d'apprentissage

1. HARenEs est sur la bonne route mais il doit être amélioré
  - L'interface doit éviter tout métalangage linguistique
  - meilleur feedback quand la recherche n'est pas bien formulée
2. Les usagers n'exploitent pas suffisamment des exemples
3. Un autre test de révision de textes pourrait donner d'autres résultats

**Bref,**

**Le CORPUS d'une façon ou d'une autre doit faire partie des ressources d'apprentissage des langues.**

**À venir, HARTA (outil pour la rédaction des textes académiques en espagnol)**

# MERCI DE VOTRE ATTENTION

margarita.alonso@udc.es

[www.dicesp.com](http://www.dicesp.com)

haren.es.taln.upf.edu/CakeHARenEs/

[www.grupolys.com](http://www.grupolys.com)

