

Autonomous collocation error correction with a data-driven approach

Orsolya Vincze

Margarita Alonso Ramos

Universidade da Coruña (Spain)

Learner Corpus Research Conference

Bergen, September 27-29, 2013

Introduction: Collocations

What is a collocation?

phraseological unit W_1W_2

W_1 = base selected according to its meaning

W_2 = collocate whose selection is determined by the base



???

pouring rain, dense fog, fierce wind

dense rain, fierce fog, pouring wind

Collocation learning and corpora

a) learner corpus → collocations are problematic

b) native corpus → collocation learning/teaching resource



Introduction: Corpora and error correction

Writing aids and NLP:

- learner writings can be checked against corpora

Benefits for L2/FL learning:

- DDL=data-driven learning
- authentic L2 input
- encourages inductive and autonomous learning

Introduction: The study

Context:

Development of an active collocation learning environment including a writing aid tool for learners of Spanish

Preliminaries:

Learner corpus derived typology of collocation errors (Alonso Ramos et al. 2010)

Aim:

Can learners autonomously correct collocation errors with the help of concordance lines? To what extent?

Outline

1. The study

2.1. Collocation error types

2.2. Research questions

2.3. Methodology

2. Results

3.1. General findings

3.2. Correction of specific error types

3.3. Evaluation of concordance lines as feedback

3.4. Enhancing concordance line feedback

3. Conclusions and future work

1. The study

1.1. Collocation error types

- Typology of collocation errors based on CEDEL2 (Lozano 2009)

1) **Lexical collocation errors**, e.g.:

Incorrect collocate: **capturar la atención* instead of e.g. *captar la atención* ‘catch sb’s attention’

Synthesis: **misinterpretaciones* instead of e.g. *malas interpretaciones* ‘wrong interpretations’

2) **Grammatical collocation errors**, e.g.:

Governed preposition: **montar una bicicleta* instead of *montar en una bicicleta* ‘ride a bike’

Number: **dimos bienvenidas* lit. ‘we gave welcomes’ instead of *dimos la bienvenida* ‘we gave a welcome’

1.2. Research questions

- 1) Can learners autonomously correct collocation errors with the help of concordance lines?
- 2) The correction of what error types poses more difficulty for the students when presented with the concordance lines?
- 3) What problems can learners have when dealing with concordance feedback?
- 4) How can concordance line feedback be improved in order to better assist students in the revision of collocation errors?

1.3. Methodology

Questionnaire

- 20 sentences from CEDEL2 (Lozano 2009) containing a collocation error
- Concordance lines: a) full sentences from esTenTen (Kilgarriff et al. 2004); b) Google Books n-grams

Tasks

- 1) propose a correction without any aid
- 2) propose a correction with the help of concordance lines

Participants

18 Spanish as a second language students working or studying in Spain at the time of the test

1.3. Methodology

11. La hija está tratando de capturar la atención de su madre, pero es muy difícil porque la madre siempre trabaja.

La ingesta de líquidos es a lo que habitualmente prestamos menos atención.

Los niños sienten celos y los expresan a través de necesidades a las que los padres deben prestar atención.

Esta siempre ha sido la enfermedad que más me ha asombrado y que me ha llamado la atención de una forma especial.

Aquí os vamos a describir algunas opciones que nos han llamado la atención por sus cualidades.

Connelly atrae la atención y no deja que decaiga en ningún momento.

El otro día estaba comprándome libros en la librería y este me atrajo la atención.

Desarrollar una cierta habilidad para el coqueteo te puede ayudar a aprender a captar la atención e interés de las personas que te agradan.

Los niños suelen portarse mal para captar la atención de los padres.

Sample questionnaire item with full-sentence concordances

11. La hija está tratando de capturar la atención de su madre, pero es muy difícil porque la madre siempre trabaja.

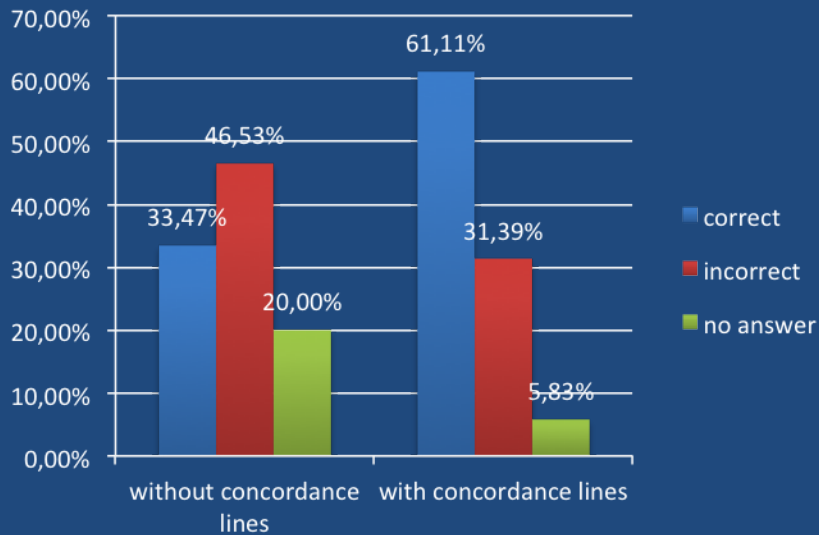
captan la atención de
captar la atención y la
para captar la atención de
concentrado la atención de los
concentra la atención de
de concentrar la atención de
captan la atención de los
concitan la atención de los
concitaban la atención de los
captan la atención de los

Sample questionnaire item with n-gram concordances

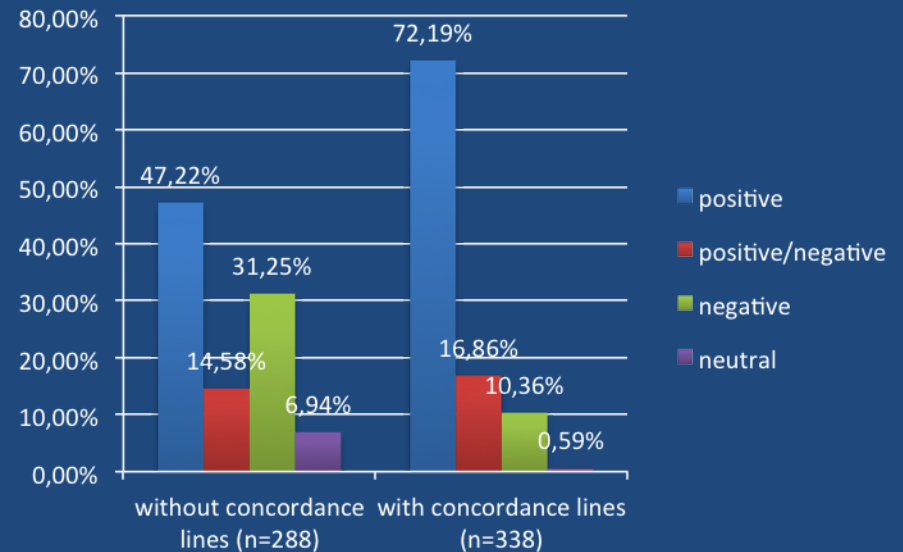
2. Results

2.1. General findings

- with concordance lines higher number of correct suggestions, while the number of incorrect suggestions, as well as questionnaire items left blank was lower
- more positive and positive/negative changes and less negative and neutral or irrelevant changes with concordance lines



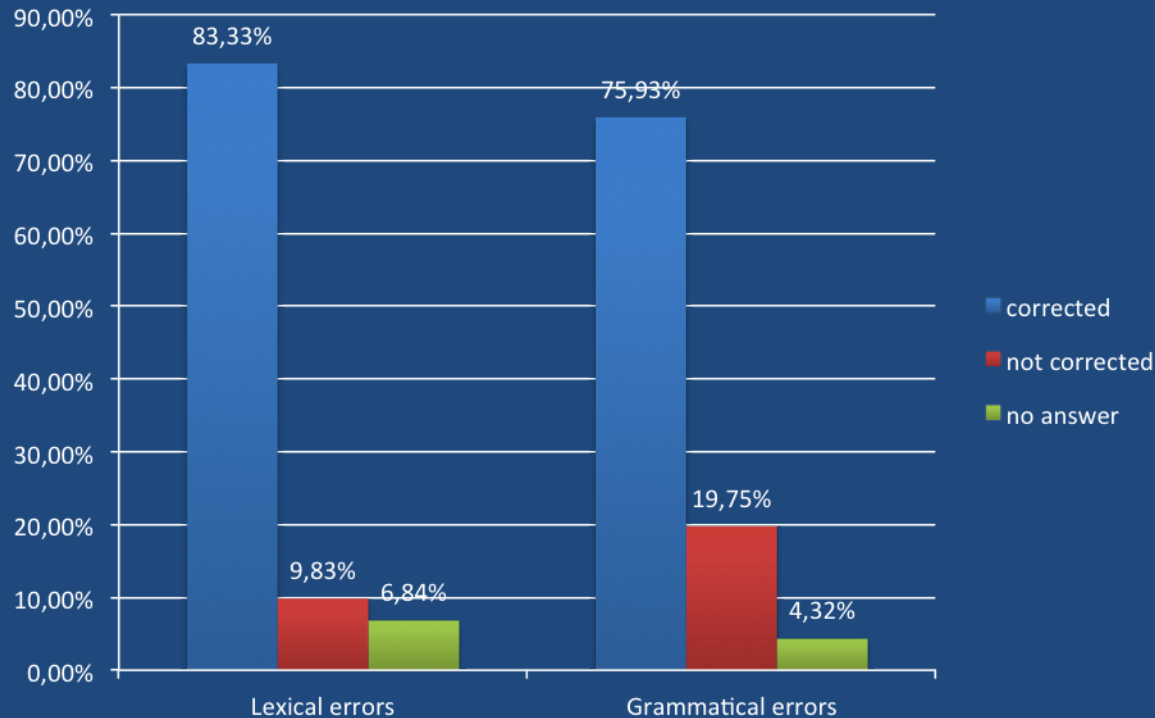
Total number of correct and incorrect suggestions, no answers provided or repeated answers (n=360)



Number of positive, positive/negative, negative and neutral changes

2.2. Correction of specific error types

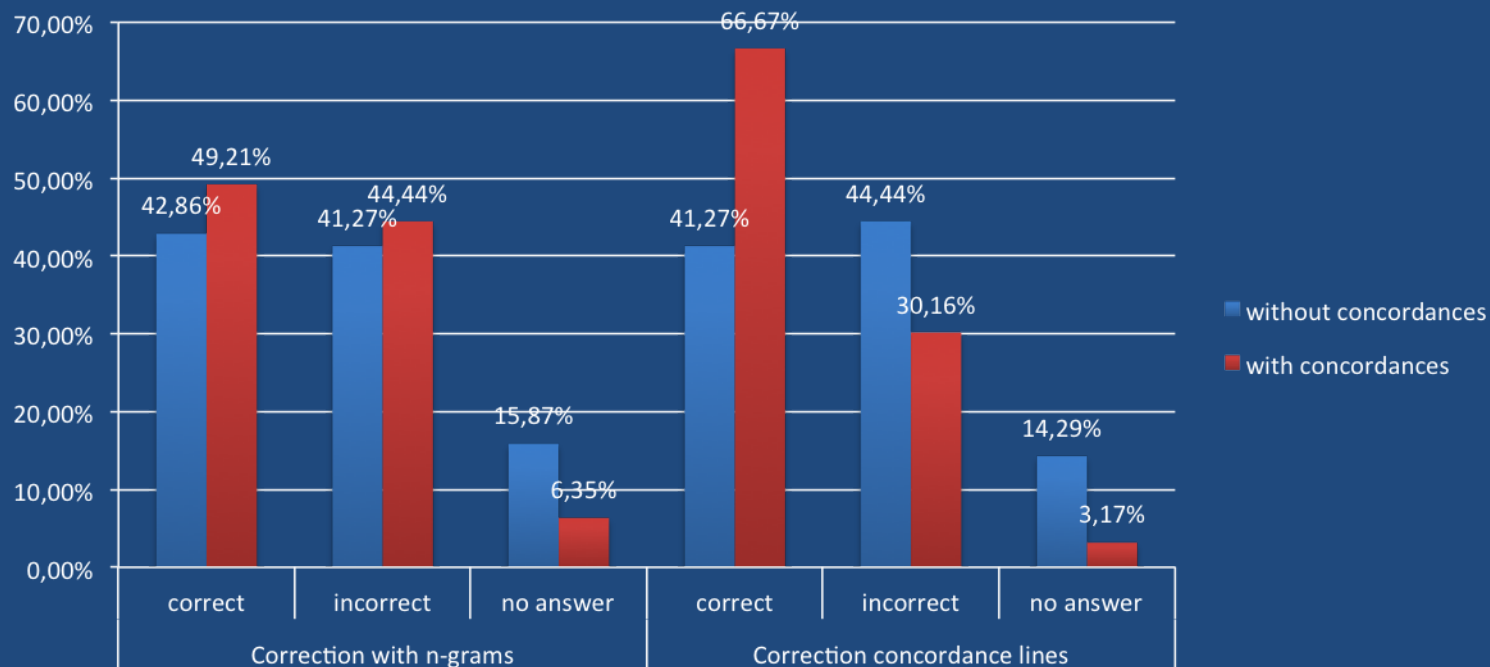
- Participants were more successful in correcting lexical collocation errors than grammatical errors
- More difficulty in noticing grammatical features



Participants' success in correcting different collocation error types with the help of concordance lines

2.3. Problems with concordance lines as feedback: comparing full-sentence and n-gram concordances

- *full-sentence concordance lines are more effective: higher number of correct and lower number of incorrect answers*



Number of correct, incorrect suggestions, no answers provided or repeated answers according to concordance type

2.3. Problems with concordance lines as feedback: Analyzing incorrect suggestions

1) New errors in participants' answers

a) non-concordance induced errors

Erroneous segment in original learner sentence	Erroneous correction suggestion	Expected correction
<p>..nos despedimos, y *<u>gracias</u>, y caminamos hacia el puerto..</p> <p>lit. 'we said goodbye, <u>thanks</u>, and we started to walk towards the port'</p>	<p>..nos despedimos, y *<u>gracias a Dios</u> caminamos hacia el puerto..</p> <p>'we said goodbye, and <u>thank God</u> we started to walk towards the port'</p>	<p>..nos despedimos, y <u>les dimos gracias</u>, y caminamos hacia el puerto..</p> <p>'we said goodbye, and <u>thanked them</u>, and started to walk towards the port'</p>

2.3. Problems with concordance lines as feedback: Analyzing incorrect suggestions

1) New errors in participants' answers

- a) non-concordance induced errors
- b) meaning-related concordance-induced errors: probably due to lack of sufficient context

Erroneous segment in original learner sentence	Erroneous correction suggestion	Expected correction
Mi futuro no <i>*tiene limitades</i> .	Mi futuro no <i>*tiene limitaciones</i> .	Mi futuro no <i>tiene límites</i> .
‘My future has no <u>limits</u> .’	lit. ‘My future <u>has no limitations</u> .’	lit. ‘My future <u>has no limits</u> .’

2.3. Problems with concordance lines as feedback: Analyzing incorrect suggestions

1) New errors in participants' answers

- a) non-concordance induced errors
- b) meaning-related concordance-induced errors: probably due to lack of sufficient context
- c) concordance-induced errors involving the inappropriate application of a pattern observed in the concordances

Erroneous segment in original learner sentence	Erroneous correction suggestion	Expected correction
* <u>La película se trata de</u> una mujer soltera, su hija y sus amigas...	* <u>La película, que se trata de</u> una mujer soltera, su hija y sus amigas..	<u>La película trata de</u> una mujer soltera, su hija y sus amigas..
‘ <u>The film is about</u> a single woman, her daughter and her friends..’	‘ <u>The film, which is about</u> a single woman, her daughter and her friends..’	‘ <u>The film is about</u> a single woman, her daughter and her friends..’

2.3. Problems with concordance lines as feedback: Analyzing incorrect suggestions

1) Negative changes in participants' answers

- a) non-concordance induced errors
- b) meaning-related concordance-induced errors: probably due to lack of sufficient context
- c) concordance-induced errors involving the inappropriate application of a pattern observed in the concordances

2) Incomplete correction of learner sentences

Erroneous segment in original learner sentence	Erroneous correction suggestion	Expected correction
...y entonces * <u>encendió el fuego que quemó la casa</u> y los mató.	...y entonces * <u>prendió el fuego que quemó la casa</u>y entonces <u>prendió fuego a la casa</u> ...
lit. '... and then <u>she lit the fire that burnt the house</u> '	lit. '... and then <u>she set the fire that burnt the house</u> ...'	lit. '... and then <u>she set fire to the house</u> '

3.4. Enhancing concordance line feedback

- **grammatical errors are less salient in concordance lines**
 - group concordance lines in order to emphasize patterns (similar to Wu et al. 2010)
- **implicit nature of concordance feedback**
 - Pro: promotes inductive learning
 - Con: participants do not always manage to identify the errors
 - should there be an explicit indication of errors?
 - multiple-step feedback: 1) only concordances, 2) additional aid
- **concordance-induced errors due to lack of context**
 - allow users to check wider context and more corpus examples if needed
- **students might need information not provided by concordance lines**
 - integration with dictionary (meaning related errors)
 - incorporate information on verb conjugation

3. Conclusions and future work

3. Conclusions

1) Can learners autonomously correct collocation errors with the help of concordance lines?

Yes, our study shows that concordance lines do have a favorable effect on learners' autonomous error correction

2) The correction of what error types poses more difficulty for the students when presented with the concordance lines?

Grammatical collocation errors are less salient than lexical collocation errors.

3. Conclusions

3) What problems can learners have when dealing with concordance feedback?

- lack of context: identify/distinguish meanings
- identification/noticing of error

4) How can concordance line feedback be improved in order to better assist students in the revision of collocation errors?

- emphasize grammatical patterns in the presentation of concordances
- allow more context
- more explicit indication of error (optional)
- integration with other language learning resources

3. Future work

HAREnEs prototype interface

Herramienta de ayuda a la redacción en español = Spanish writing aid tool

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

Es | En

taln DiCE

Inicio Comprobar colocaciones Buscar colocaciones Mi diccionario Acerca de Accede:

Bienvenido a HAREnES, la Herramienta de Ayuda a la Redacción en Español.

Uno de los problemas más recurrentes a los que se enfrenta un aprendiz a la hora de escribir textos en una lengua extranjera es el uso de las colocaciones. Las colocaciones son combinaciones de palabras restringidas, es decir, que no se combinan de forma libre. Por ejemplo, en español decimos *alegría loca* para expresar que la alegría es 'muy grande', pero no decimos **vergüenza loca*, sino *vergüenza terrible*.

Esta herramienta permitirá al aprendiz corregir y buscar colocaciones en español, así como crear su propio diccionario de colocaciones.

La interfaz HAREnES ha sido producida con el fin de poner los avances de un proyecto de investigación en curso al alcance del usuario. Esto que significa que la interfaz es necesariamente de un carácter experimental.

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.

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.

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.

Búscanos en Facebook

DiCE Diccionario de Colocaciones del Español

Me gusta Te gusta esto.

DiCE Diccionario de Colocaciones del Español

Ayer encontramos este ocurente artículo de J.J. Millás que os invitamos a leer sobre el diccionario Redes, a propósito del beso que la reina Sofía dio al monarca. Un beso puede ser "apasonado".

COLOCATE

DIXIO

v/* (behavior:urk/#default#VML);) ot*
(behavior:urk/#default#VML);) wt*
(behavior:urk/#default#VML);) abape
(behavior:urk/#default#VML);) ...

La competencia en...

La II edición del Concurso de blogs de promoción del español y la cultura española nos ha permitido conocer otros blogs sobre la enseñanza del espa...

Writing Assistant

Hemos descubierto una nueva página a través de Facebook llamada Writing Assistant que queremos compartir con vosotros. Se trata de una herramienta ...

No eres tú... son t...
Normal 0 21 false false false ES X-NONE X-NONE ...

Feed/Feed

3. Future work

HAReNs prototype interface

Herramienta de ayuda a la redacción en español = Spanish writing aid tool

The screenshot shows the HAReNs web interface. At the top, there is a yellow header with the title 'HAReNs' and a navigation menu with items: 'Inicio', 'Comprobar colocaciones', 'Buscar colocaciones', 'Mi diccionario', and 'Acerca de'. The main content area is titled 'Comprobar colocación' and contains the instruction: 'Introduce una colocación para comprobar si es correcta (p.ej. dar un paseo o tomar un paseo)'. Below this, the input field contains 'tomar un paseo' with a red 'X' icon indicating it is incorrect. A yellow 'Comprobar' button is visible. The result section states: 'Esta colocación es incorrecta, te mostramos algunas sugerencias de corrección:'. A table of suggestions follows, with a 'Ver ejemplos' link and a red 'X' icon. The table has three columns: the first column shows the original phrase with a bracket indicating the location of the error, the second column shows the suggested correction, and the third column shows an example sentence with a green checkmark icon.

Original phrase	Suggestion	Example
...ta se baja del buque para]	dar un paseo	por mar pudiéndose baña...
...dad , e incluso se permite	dar un paseo	aéreo a su enamorada , l...
...uerpo , juvenil de espíritu	daba un largo paseo	por las murallas .
fácil ver Brujas , Gante y]	dar un paseo	por Bruselas , donde la vi...
se vestía y se iba a]	dar un paseo	por las tierras andaluzas ...
...ilidad de distracción para]	dar un paseo	, tomarse un café , o conv...
... del todo , así decidimos]	dar un paseo	ligero sin alejarnos much...
...café para despejarnos y]	dar un largo paseo	por la ciudad semi dormid...
...viven cómodamente para]	dar un paseo	por la ciudad .]
... acabar Ayer noche salí a	dar un paseo	por la céntrica zona de C...
...e 2004) Hoy he salido a]	dar un paseo	en esta tarde de invierno , ...
pues el sábado fuimos a]	dar un paseo	ya el tiempo nos concedí...
...las trompetas se fueron a]	dar un paseo	por la playa .]
... mientras Serrano y -yo »	dábamos un paseo	-tiró sobre nosotros dema...
...obre dos ruedas , porque	dando un paseo	o yendo en el coche , con...
fui a]	dar un paseo	muy largo para refrescar...

Thank you for your attention!

This research has been supported by Ministerio de Economía y Competitividad (FFI2011-30219-C02-01), and the FPU grant (AP2010-4334).