Falling into the realization over time: A diachronic study of *caer en la cuenta*

Aubrey Healey
University of New Mexico

ABSTRACT. This study examines the evolution of the Spanish phrase *caer en la cuenta* 'to realize/be aware' from its first occurrence in the 1500s through the 1900s using the Mark Davies Corpus de Español. This research was inspired by a previous investigation (Healey 2012) into the changes that took place in another linguistic construction, *darse cuenta*, with the same meaning. The present study finds that *caer en la cuenta*, doesn't appear in the corpus till several centuries after the *darse cuenta* construction has already been in existence, and not till after the *darse cuenta* developed the 'realize/be aware' meaning. Additionally, *caer en la cuenta* only ever occurs with the 'realize/be aware' meaning. These findings, and the similar syntactic changes both constructions undergo, lead to the conclusion that *caer en la cuenta* has lexicalized and developed as a result of analogy with the *darse cuenta* construction.

Keywords: Language change, lexicalization, analogy, Spanish, diachronic, corpus-based

1. INTRODUCTION. By analyzing the semantic evolution of the construction [caer]+*en*+(*la*)+*cuenta* from the 1500s to the 1900s, this pilot study demonstrates that a particular semantic construal has developed for this construction based on the aggregate context of use and from analogy with the *darse cuenta* construction. The possible semantic readings of this construction over the last five centuries were all related to cognitive action, with the most common interpretations being 'understand', 'be aware', and 'realize.' Since these interpretations can all be seen as the same kind of mental action with the only difference being aspect, it was realized that all tokens really had the same meaning. This is different than the *darse cuenta* construction, which will be discussed in what follows. These examples show an early and late stage of the *caer en la cuenta* construction.

(1) Y entiendo cierto que ellos mismos, cuando se hallen reducidos y virtuosos,

han de caer en la cuenta de la perdición que tenían, y quedar muy

contentos y agradecidos de tanto bien como se les ha hecho; que tiene tanta

fuerza la verdad y virtud como esto.

'And I certainly understand that they themselves, when they find themselves

restricted and moral, **they have to realize the damnation they used to have**,

and be very happy and grateful for all the good that's been done for them;

that truth and righteousness have as much strength as this.'

(CdE, Pérez de Herrera, Cristóbal, *Discurso de amparo de los legítimos*

pobres y reducción de los fingidos, 1500s)

(2) Trato entonces de reconstruir mentalmente el mobiliario de cada habitación y

caigo en la cuenta de que ese mueble podría ser el de Carlitos y Arturo.

'I try then to mentally reconstruct the interior design of each room and **I**

realize that that furniture could be Carlitos and Arturo's.'

(CdE, Simón, Carmen, *Cubo de Luz*, 1900s)

As can be seen in the above examples, the 'realize/be aware/understand' meaning has occurred throughout the construction's existence. The construction's occurrence with that meaning throughout its history was not the original hypothesis of the author. It was expected that the construction would have gone through periods of ambiguity like many evolving constructions (e.g. the development of English *going to* as a future marker). Possible reasons for this different evolutionary path will be discussed in the body of this article. A recent study of the *darse cuenta* construction (Healey 2012) will be referenced for comparison in this article, as it informs the current study's methodology and interpretation.

2. THEORETICAL FRAMEWORK. This section provides a cursory literature review of the theories necessary for understanding the approach to this investigation and interpreting the results. The theories and approaches called upon in this paper, discussed below, are: construction grammar (Croft 2001, Goldberg 2006, *inter alia*), lexicalization (Brinton & Traugott 2005, Himmelmann 2004, Lehmann 2002, *inter alia*), analogy (Hilpert 2013, Itkonen 2005), and exemplar theory (Bybee 2013, Pierrehumbert 2001).

2.1. CONSTRUCTION GRAMMAR. There are a variety of approaches to construction grammar, but they all share commonalities that unite the theory as one. The most important tenet, shared by construction grammar ideologies, is that language is made up of constructions, which are form-meaning pairings. These form-meaning pairings are the most primitive units of meaning. The form part of the pair includes morphosyntactic, lexical and phonological properties of the construction, and the meaning part includes its semantic, pragmatic, and discourse-functional properties. Because of this, in general, construction grammarians do not presuppose a division between lexicon and syntax. Instead, all constructions form part of a continuum from more to less schematic meanings. For example, the ditransitive *give* construction [give] + RECIPIENT + THEME (Malchukov et al. 2010) construction has a very schematic meaning with several open slots. The compound noun *fundraiser* (also a construction), on the other hand, is not schematic at all. It is completely lexically-filled, and the pattern is not productive, meaning that it can only be used in this one form with this particular meaning.

Categorization drives the mental organization of language in the mind of a speaker, and can take place at many levels of abstraction (Langacker 1987, Wilson 2014). The abovementioned continuum of schematicity is represented in the mind of a speaker as a network of constructions, and is shaped by the repeated exposure to specific utterances and domain-general cognitive processes such as categorization and cross-modal association (Hoffmann & Trousdale 2013:4). This study takes these principles as fundamental to the way analysis of the data is carried out and interpreted.

2.2. LEXICALIZATION. This study suggests that the *caer en la cuenta* construction has undergone lexicalization within the speech community. Brinton & Traugott (2005:96) define lexicalization as:

> A change where in certain linguistic contexts speakers use a syntactic construction or word formation as a new contentful form with formal and semantic properties not predictable from the constituents of the construction or the word formation pattern. Over time there may be further loss of internal constituency and the item may become more lexical.

Certain implications follow from this definition (Brinton & Traugott 2005:96-97), namely: (1) the result of this process is a new contentful form-meaning pairing (Himmelmann 2004), (2) the input to lexicalization can be a form-meaning pairing of any semantic or formal complexity (although in the view of Lehmann 2002 only complex units can lexicalize), (3) the same applies to the output, (4) additional morphological, phonological, and semantic changes may modify existing forms, (5) lexicalization is a gradual change (Trousdale 2008, Traugott 2012), (6) fusion/univerbation typically occur (Himmelmann 2004, Lehmann 2002), (7) idiomatization/demotivation tend to occur (where the semantic parts of the construction lose compositionality, e.g. *never*, *the*, and *less* in *nevertheless* don't mean what they would separately or even combine the way they would in basic syntax) (Bybee & Torres Cacoullos 2009, Lehmann 2002, Trousdale 2008), and (8) there is typically a decrease in pattern productivity (Himmelmann 2002).

A prototypical example of lexicalization is the development of the word *barn* from the combination of Old English words *bere* 'barley' and *ærn* 'house'(example from Brinton & Traugott 2005:97). First, speakers started to use the combination *bere ærn* to describe where they

kept their barley, then (gradually) these words became fixed in this order without intervening

constituents, then they became morphologically fused together, and eventually the unit became

phonologically eroded as the originally distinct words became demotivated. The lexeme +

lexeme combination became just one lexeme. Semantic reanalysis occurred (first) to make

speakers want to conceptualize this barley house as one construal, a separate idea than just a type

of house. By the time speakers are calling this item a *barn*, it has generalized to be more than just

a house for barley; it has become a place to store hay and tools as well. It is also possible that

speakers named that concept 'barley house' in a metonymic fashion, even though they kept more

than barley in it from the beginning. Some items which have been considered cases of

lexicalization in the sense of the above definition are: fused syntactic phrases accompanied by

idiomatization and sometimes undergoing morphosyntactic change, fused compounds,

phonogenesis (the creation of new syntagmatic phonological segments out of old morphemes, e.g.

while-s- t 'while + GEN + EXCRESCENT' > *whilst* 'during', *ibid*), phonologization (when a

phonetic difference becomes a meaningful difference between phonemes, e.g. *drink*, *drench*),

and the creation of semantic non-category changing affixes (Brinton & Traugott 2005:98).

2.3. ANALOGY. Analogy is another important process to this research because it is posited that

the *caer en la cuenta* construction resulted from analogy with an earlier construction (*darse*

cuenta) that has the same meaning and a very similar form. According to Itkonen (2005),

analogy is generally defined as 'structural similarity,' meaning that both the wholes and the parts

of items generally match up. For example, *off the hook*, meaning something is exciting and fun,

is analogically related to *off the chain* which has the same meaning and a very similar form.

Hilpert (2013) operationalizes analogy as "an increasing range of elements that occur within a

given constructional slot." The constructional slot which allows more elements over time in this study is the verb change between _darse_ *cuenta* and _caer_ *en la cuenta*, which will be discussed further in section 5.2.

2.4. USAGE-BASED EXEMPLAR THEORY. Humans are great categorizers, and our individual knowledge networks are organized in terms of exemplars based on similarity in form and meaning (Bybee 2003). An exemplar is a category which emerges from tokens of experience that are judged to be the same (Bybee & Beckner 2010, Bybee 2013). The one assumption that makes this understanding possible is that language emerges through use (Hopper 1998). The meanings of constructions are influenced by their surrounding discourse context and the actual situation in which they are used (Blank 2001, Heine 2002, Himmelmann 2004). The contexts of use contribute nuances and connotations to the meaning of a construction. A form-meaning pairing will become more entrenched with repeated use. Each particular token of use is represented (although these memories and pathways decay if not reinforced with repeated use). Therefore, even the individual token of, c*aí en la cuenta de que no fue al mercado* ('I realized that s/he didn't go to the grocery store'), makes a mark on the representation of the *caer en la cuenta* construction. Repeated exposure to *caer en la cuenta* with the meaning 'realize/be aware/understand,' lexicalizes that string as a set way to communicate that idea.

3. DATA. Two hundred and ninety-eight tokens of [caer]+*en*+(*la*)+*cuenta* from the 1500s to the 1900s were extracted exhaustively from the Mark Davies Corpus del Español (Davies 2002). This searchable online corpus contains 101,311,682 words from 13,926 different texts from the 1200s to the 1900s. All of the texts in the corpus come from written discourse except for around

2,040 oral interview transcripts in the 1900s (5,113,249 words). The same corpus is used for this study as was used in an earlier study of *darse cuenta* (Healey 2012) for comparability. The *caer en la cuenta* construction does not appear until the 1500s, although the corpus itself contains records all the way back to the 1200s.

4. METHODOLOGY. Exhaustive extraction of tokens was undertaken. The search method used was the lemma [caer], with *cuenta* within 5 words to the right of the verb in a single clause. This search method was deemed better than searching for the actual string *caer en la cuenta* because this way the tokens that had intervening constituents would be discovered as well. The tokens were then compiled in EXCEL with their surrounding discourse context and source information. Each token was then coded for the following factors: (1) semantic reading of the construction, (2) century, (3) complement type, (4) whether or not the noun phrase (NP) complement, if it has one, represents an event, and (5) the presence or absence of intervening constituents. Pivot tables in EXCEL were used for comparison, but no statistical tests were run due to the relatively small sample size, and the homogenous semantics of the tokens.

The reason for coding the semantics and century of occurrence for each token, is fairly obvious. We are able to see if any changes in speakers' interpretation of the phrase took place over time, and also when those changes took hold. Unfortunately for these data, no semantic change took place, so the results for that factor were not very interesting, but informative nonetheless. The third factor, complement type, was coded for based on the findings of a previous study on *darse cuenta* (Healey 2012). It was discovered that there is an increase in the variety of complements that the phrase took over time, and specifically an increase in noun phrase (NP) complements and predicate complements over time. This discovery demonstrates

that the phrase is being used more and more with the discourse function of introducing *what* someone is realizing, and also that the way speakers describe what they are realizing is accomplished in different syntactic ways at different points in time. Of course, multiple complement types occur with the construction at any point in time due to natural variation. The *darse cuenta* construction took complement types such as: noun phrase (e.g. *ellos* 'them,' *lo* 'it,' *su virtud* 'her virtue'), predicate (e.g. *que él está llorando* 'that he is crying'), *si* 'if' clause (e.g. *si va a venir o no* 'if she's coming or not'), relativizer (e.g. *lo que* 'it that/which'), and infinitive (e.g. *llegar* 'to arrive/arriving'). It was later decided that *si* 'if' clauses are really a type of predicate, so those categories were collapsed for the present study. Note that an infinitive is used in some cases where the English translation would use a gerund.

Another finding from the *darse cuenta* study (Healey 2012) was that within the tokens with a noun phrase complement there was an increase in the number of NP complements whose semantics represented events or more abstract ideas. An example of an event NP is given below.

(3) **No se dieron cuenta de** <u>nuestra llegada</u>; por esto no huyeron hasta que

   estuvimos próximos a su poblado y no pudieron escapar.

   '**They were not aware of** <u>our arrival</u>; because of this they didn't flee until we

   were close to their village and they couldn't escape.'

   (CdE, Schmidel, Ulrico, *Relatos de la conquista del Río de la Plata y*

   Paraguay: 1534-1554, 1500s)

The increase in abstract and event NPs was an important finding because a referent that semantically represents an action or intangible idea can be seen as semantically having the same

function as a predicate. It is just a matter of different ways of expressing something hard to describe, an abstract idea or something that happened. The abstract and event NPs are a bridging context for the *darse cuenta* construction as its function changes over time and the syntax catches up to match it. The last factor I coded for was the presence or absence of intervening constituents, that is, whether or not there are extra words in the phrase. It is expected that when a phrase is new it will be less fixed in form. Thus, as time goes on there should be less and less extra words intervening in the phrase.

After all of the tokens were coded for the above factors in EXCEL, pivot tables were used to discover tendencies for each factor as related to semantics and time. No statistical tests were run on these data because of the small sample size (298 tokens) and the uniformity in the dependent variable (semantics). Additionally, this is a pilot study for a larger project, so the level of investigation is not as deep as it could be.

5. RESULTS AND DISCUSSION. One of the major findings of this research is that the token frequency of NP complements that occur with [caer]+*en*+(*la*)+*cuenta* decreases over time, and the frequency of predicate complements increases. This shows that [caer]+*en*+(*la*)+*cuenta* is increasingly fulfilling the discourse function of introducing another clause, the clause that explains what someone is realizing, much like [PRO]+[dar]+*cuenta*. The [caer]+*en*+(*la*)+*cuenta* construction seems to have lexicalized with this discourse function. That is, the construction's form is fixed and meaning is constant. The data also reveal that the [caer]+*en*+(*la*)+*cuenta* construction, which does not appear in this corpus until the 1500s, patterns similarly to [PRO]+[dar]+*cuenta* from the 1500s onward in both semantics and the complement types it accepts. Since the [caer]+*en*+*la*+*cuenta* construction appears in the corpus at a later date, it is

posited that *caer en la cuenta* resulted via analogy with *darse cuenta*. In addition, it seems that [PRO]+[dar]+*cuenta* might be the type-leader for the pattern of Verb + *cuenta* with a semantic reading of cognitive action ('realizing') since the *darse cuenta* construction has been around longer and is more frequent than the *caer en la cuenta* construction. This is a question to be investigated further in a joint study of both constructions. The following two subsections will discuss how the results of the coding demonstrate that *caer en la cuenta* has lexicalized and why it seems to be a result of analogy with *darse cuenta*.

5.1. LEXICALIZATION. The most subtle, yet telling, finding of this study is that the *caer en la cuenta* tokens have relatively unvarying semantics over time. All of the tokens can be understood as 'realize/be aware/understand,' although it is possible to see some nuanced differences between tokens such as the tiny semantic difference between 'figure out' and 'discover.' Conventionalization of a phrase occurs when all the members of the speech community agree as to what the meaning of a particular construction is. This phrase has been conventionalized as a way to say 'realize' since all Spanish speakers can understand it that way. Any of the nuanced differences are suggested by the greater discourse context. Syntactically, there is a decrease in the occurrence of intervening constituents over time, thus the phrase is becoming fixed in form. The percentage of *caer en la cuenta* tokens occurring with intervening constituents in each century is given in table 1 below.

Century	1500s	1600s	1700s	1800s	1900s
%	17.5%	8.1%	15.3%	0.9%	6%
n intervening/total n for that century	14/80	5/61	4/26	1/106	3/50

TABLE 1. Percent of intervening constituents by century

There are in general two types of intervening constituents: subjects, because of the relatively free order of subject-verb in Spanish and temporal adverbs (e.g. *luego* 'later,' *ahora* 'now,' *tan tarde* 'so late,' *ya* 'already'). They are pieces that add slightly more information to the picture of 'realizing,' for example *when* someone realized something.

There is also a decrease in the number of complement types over time, skewing towards predicates, and to a lesser extent NPs. As can be seen in table 2 below, when the phrase does take a complement it is usually a predicate or a noun phrase (NP). Up until the 1900s, other complement types were possible, such as infinitives (e.g. *necesitar* 'to need'), relativizers (e.g. *lo que* 'it that'/'which'), and even one example of a comparative (*mejor* 'better') used in a less idiomatic fashion as in 4 below.

(4) y han de caer en la cuenta mejor que en el yerro, y conocer lo que es bien y

mal.

'and they have to fall on the account better than on the error, and know what

is good and bad.'

(CdE, Espinel, Vicente *Vida del escudero Marcos de Obregón*, 1500s)

The fact that the variety of complement types decreases over time shows that the way speakers express what they are realizing is becoming more conventionalized.

Complement	Inf.	NP	Relativizer	Predicate	Comparative	None	Total
Century							
1500s	2	26	8	24	0	20	80
1600s	0	11	5	15	1	29	61
1700s	0	4	1	11	0	10	26
1800s	3	19	6	56	0	23	106
1900s	0	11	0	35	0	4	50

TABLE 2. Complement types by century

Based on the decrease in the frequency of intervening constituents over time (fixation in form), the reduced variation in the complement types that the phrase occurs with (fixation in function), and the relatively unvarying semantics from token to token, it seems that these constructions are becoming specialized in function and fixed in form. The fixation in form and meaning demonstrates that this phrase has lexicalized.

5.2. ANALOGY. There are two reasons that the *caer en la cuenta* construction appears to have resulted from analogy with the *darse cuenta* construction. Firstly, the *caer en la cuenta* construction does not even appear in the Davies corpus until the 1500s. The *darse cuenta* construction first appears with the 'be aware/realize' construal in the 1500s. The extension from 'give someone an account' *dar**le** cuenta* to 'give yourself the account' ('realize') *dar**se** cuenta* is a logical path for the construction; syntactically, notice the change from the dative pronoun *le* to the reflexive pronoun *se*. However, *caer en la cuenta* appears in the 1500s with no apparent semantic bridging context, only the example of *darse cuenta*'s new meaning to follow. In the *darse cuenta* study (Healey 2012) it was posited that the abstract noun phrase and event nominalization complements were a bridging context that allowed speakers to imagine their "realizations" as predications. The NP complements that *caer en la cuenta* took during the

1500s-1900s were events or abstract ideas more often than concrete noun phrases consistently over time. Essentially, *caer en la cuenta* made its first appearance with all of the features that *darse cuenta* evolved to have, which is why it is a clear example of analogy.

6. CONCLUSION. In sum, the data reveal that the construction [caer]+*en+(la)+cuenta* firstly, appeared as a result of analogy with the extant construction *REFL+*[dar]+*cuenta,* and secondly, has lexicalized. The case for analogy with *darse cuenta* is evidenced by the fact that the *darse cuenta* construction existed for several centuries before the *caer en la cuenta* construction even appeared in the corpus, and also by the fact that when *caer en la cuenta* did appear, the *darse cuenta* construction had already developed the 'realize/be aware' meaning. *Darse cuenta* had a path of development that follows logically from the changing contexts in which it was used over time. *Caer en la cuenta,* on the other hand, simply appeared with this idiomatic meaning.

There are three pieces of evidence for the lexicalization of *caer en la cuenta:* (1) the relative homogeneity in meaning, (2) the decrease in complement types over time, and (3) the decrease in the number of intervening constituents over time. As mentioned earlier, all of the tokens in the corpus can be understood to mean something like 'realize/be aware/understand' with only slight variations in exact meaning due to the greater discourse context. This is evidence for lexicalization because one of the ways you can tell if something has lexicalized is if there is no variation in meaning. The other part of lexicalization is fixation in form, which is demonstrated by the decrease in the occurrence of intervening constituents over time and the decrease in the variety of complement types that occur with it over time. If less words can interrupt the phrase, it is understood by the speakers more as a single unit. The fact that only two complement types (noun phrases and predicates) can occur with the phrase by the 1900s shows

that the way speakers are expressing what they have "realized" has conventionalized as well. Thus, the *caer en la cuenta* construction has both lexicalized and is the result of analogy with the earlier *darse cuenta* construction.

REFERENCES

BYBEE, JOAN L. 2013. Usage-based theory and exemplar representations of constructions. In Thomas Hoffmann & Graeme Trousdale, *The Oxford Handbook of Construction grammar*, Oxford: Oxford University Press. 49-69.

BYBEE, JOAN AND CLAY BECKNER. 2010. In Bernd Heine and Heiko Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press. 827-856.

BYBEE, JOAN AND RENA TORRES CACOULLOS. 2009. The role of prefabs in grammaticization. In Roberta Corrigan, Edith Moravcsik, Hamid Ouali, and Kathleen Wheatley. *Formulaic Language*. Vol. 1, Distribution and historical change, Amsterdam: John Benjamins. 187-217.

CROFT, WILLIAM. 2001. *Radical Construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.

DAVIES, MARK. 2002-present. Corpus del Español (100 million words, 1200s-1900s). Available online at http://www.corpusdelespanol.org. Accessed on 05/30/2014.

FILLMORE, CHARLES J., PAUL KAY, AND MARY CATHERINE O'CONNOR. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 64: 501-538.

GOLDBERG, ADELE. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

HEALEY, AUBREY. 2012. From giving to epiphany: A diachronic analysis of *darse cuenta*. Presented at *Simposio de Literatura y Lingüística del Departamento de español y portugués*. University of New Mexico.

HEINE, BERND AND HEIKO NARROG (Eds.). 2010. *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press.

HILPERT, MARTIN. 2013. Corpus-based approaches to constructional change. In *The Oxford Handbook of Construction Grammar*, Thomas Hoffmann & Graeme Trousdale (Eds.). Oxford: Oxford University Press. 458-475.

HIMMELMANN, NIKOLAUS. 2004. Lexicalization and grammaticalization: opposite or orthogonal? In Walter Bisang, Nikolaus Himmelmann and Björn Wiemer, *What makes grammaticalization—a look from its fringes and its components*, Berlin: Mouton de Gruyter. 19-40.

HOFFMANN, THOMAS AND GRAEME TROUSDALE (Eds.). 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

ITKONEN, ESA. 2005. *Analogy As Structure and Process: Approaches in Linguistic, Cognitive Psychology, and Philosophy of Science*. Amsterdam: John Benjamins.

LANGACKER, RONALD. 1987. *Foundations of Cognitive Grammar*. Vol. 1: *Theoretical Prerequisites*. Stanford, CA. Stanford University Press.

PIERREHUMBERT, JANET B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan L. Bybee & Paul Hopper. *Frequency and the Emergence of Linguistics Structure*, Amsterdam: John Benjamins. 137-157.

TRAUGOTT, ELIZABETH CLOSS. 1989. On the Rise of Epistemic Meanings in English: An Example of Subjectification in Semantic Change. *Language*. 65(1). 31-55. Linguistic Society of America.

TRAUGOTT, ELIZABETH CLOSS. 2012. Toward a coherent account of grammatical constructionalization. Draft for a volume on historical Construction grammar edited by Elena Smirnova, Jóhanna Barðdal, Spike Gildea and Lotte Sommerer.

Traugott, Elizabeth Closs and Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Trousdale, Graeme. 2008. Constructions in grammaticalization and lexicalization: evidence from a composite predicate in the history of English. In Graeme Trousdale and Nikolas Gisborne (Eds.) *Constructional approaches to English grammar*, . Berlin: Mouton de Gruyter. 33-67.

Wilson, Damián Vergara. 2014. *Categorization and Constructional Change in Spanish Expressions of 'Becoming.'* Koninklijke Brill NV: Leiden, The Netherlands.