

# Relative Frequency in the Grammaticization of Collocations: Nominal to Concessive *a pesar de*

Rena Torres Cacoullós  
University of New Mexico

## 1. Nominal to concessive grammaticization

In a modern Spanish example as in (1), *a pesar de* is a concessive connective. The typical form of concessives is ‘although p, q’, where both component clauses ‘p’ (*el piso tiene sesenta y cinco metros*) and ‘q’ (*parece más grande*) are entailed and there is a conflict between ‘p’ and ‘q’ (König 1985:265). On the other hand, in the 12<sup>th</sup> century example in (2), *pesar* means ‘sorrow, regret,’ and the *de* + human adnominal phrase, *del rey*, is a genitive denoting the sentient being who suffers the sorrow. Clearly, between (2) and (1) grammaticization has occurred, whereby a nominal construction has evolved into a connective. König (1985:267-68) notes that “lexicalizations of notions of conflict, obstinacy, dissonance” are a common cross-linguistic source of concessives, as with English *in spite of*, *despite* or French *en dépit de*, *au mépris de*.

- (1) Concessive: although p, q  
No, es que este piso **a pesar de** que tiene sesenta y cinco metros, parece que es más grande (XX, COREC, CCON007A.95)
- (2) Preposition *a* + noun *pesar* + genitive (sufferer)  
Por esta occasion fue preso Daniel, **a pesar del** rey que lo quiere enparar (XII, Fazienda, 179)

The data for the present diachronic study were extracted from 24 texts spanning the 12<sup>th</sup>-20<sup>th</sup> centuries. Nine are from the Old Spanish period (12<sup>th</sup>-15<sup>th</sup> century), four from the 17<sup>th</sup>, five from the 19<sup>th</sup>, and six from the 20<sup>th</sup> century, including four speech corpora (see Corpus, before References).<sup>1</sup>

This empirical study supports a view of grammaticization as the evolution of collocations into single units. In this view, morpho-syntactic decategorialization and semantic bleaching of lexical items such as *pesar* happens in COLLOCATIONS. Scholars have underscored that it is instantiations of constructions that grammaticize, not individual lexical items, for example, *be going to* rather than *go*. Grammaticizing collocations undergo increasing opacity and eventual loss of internal structure, which leads to autonomy from both their individual components and other associated constructions (Bybee 2003). Such internal fixedness goes hand in hand with external flexibility, or syntactic generalization (Traugott 2003).

The role of frequency in grammaticization, and more generally in variation and change, is receiving increased scrutiny (e.g., Bybee & Hopper 2001). Here I provide evidence that a crucial frequency measure is RELATIVE FREQUENCY, that is, the frequency of a collocation with respect to occurrences of the lexical item outside the collocation. Relative frequency is important because it promotes the autonomy of the new fused unit from its erstwhile lexical component.

---

<sup>1</sup> The primary corpus comprises (portions of) entire texts rather than the CORDE (<http://www.rae.es>) or Corpus del Español (<http://corpusdelespanol.org>), to maximize contextualization of tokens.

## 2. Decategorialization: loss of nominal trappings

In Old Spanish the infinitive nominalization *pesar* is a full-blown noun. This is illustrated in (3), where nominal properties are indicated by typical trappings, such as number (plural) and gender (masculine), determination (definite article *los*), and coordination with a non-derived noun (*batallas*). Unlike productive innovative nominalizations that name the occurrence or fact of the situation designated by the verb (such as, *aquel verle en todas partes sería casualidad* (IX, Regenta)), lexicalized nominalizations like *pesar* are infinitives only etymologically. *Pesar* is one of the earliest such lexicalized nominalizations according to Lapesa (1984:68-9). It appears with its own listing (as a noun) in the first 1726 *RAE Autoridades* dictionary as meaning ‘pena, pesadumbre, penalidad.’

(3) Agora dexamos aqui las razon<e>s delos **pesares** & delas otras batallas (XIII, GE, fol.287r)

As *pesar* increasingly occurs as part of the *a pesar de* collocation, it loses its nominal trappings. This is what Hopper (1991:22) calls decategorialization: “forms undergoing grammaticization tend to lose or neutralize the morphological markers and syntactic privileges characteristic of the full categories Noun and Verb.” Decategorialization of *pesar* may be measured by the loss of plural marking, the drop in determiners and adjectival modification, and the decline in coordination with (non-derived) nouns.

In Old Spanish texts, plural *pesares* occurrences, as in (3) above, make up 6% of all *pesar* tokens. Table 1 shows that in 17<sup>th</sup>, 19<sup>th</sup> and 20<sup>th</sup> century data, only three plurals appear in close to 500 tokens. The virtual disappearance of plural marking is one indication of degraded noun-hood.

	Singular	Plural
XII-XV	94% (187/199)	6% (12/199)
XVII-XX	99% (449/452)	1% (3/452)

Table 1: Drop in plural marking

A second measure of decategorialization is the decline in determination, depicted in Table 2. In Old Spanish, fully one-fifth, or 20%, of *pesar* tokens are preceded by the definite article. Another 44% have *gran* or another pre-nominal, such as *algún*, *mayor*, *mucho*, or *ningún*, in the determiner slot, as in (4), *a mucho descanso y sossiego, mucho pesar y tristeza*. In total, 67%, or two-thirds, of Old Spanish occurrences have some kind of determiner. In contrast, in 19<sup>th</sup> and 20<sup>th</sup> century data, only possessives occur with any appreciable frequency, as in (5), *a su pesar*. Cases of *pesar* preceded by any kind of determiner drop to 11% in the 17<sup>th</sup>, 8% in the 19<sup>th</sup>, and 3% in the 20<sup>th</sup> century data.

	Def. article	Adjective	Possessive	Other	Total (N)
XII-XV	20% (39)	44% (87)	2% (4)	2% (4)	67% (134/199)
XVII	5% (4)	0	6% (3)	2% (2)	11% (9/81)
XIX	2% (4)	1% (1)	5% (9)	1% (2)	8% (17/196)
XX	0	1% (2)	2% (4)	0	3% (6/174)

Table 2: Drop in determination

(4) Finalmente, a mucho descanso y sossiego, mucho **pesar** y tristeza (XV, Celestina, 215)

(5) Era una mueca fugaz, algo resentida; la de quien, muy a su **pesar**, se ve forzado a reconocer el talento de un adversario (XX, Tabla, 201)

A third measure of loss of noun-hood is the decrease in adjectival modification (Table 3). In the Old Spanish data, 45% of *pesar* tokens have a pre-nominal adjective, another 10% are antecedents of a relative clause, and there are also cases with post-nominal and predicate adjectives. These uses are illustrated in (6)-(9): pre-nominal adjective *improviso pesar*, post-nominal adjective *pesar durable*, predicate adjective *el pesar fuese grant*, and relative clause *el pesar que estos locos me fizieron*. In

striking contrast, in 17<sup>th</sup>, 19<sup>th</sup> and 20<sup>th</sup> century data, only 12 of nearly 500 tokens (3%) have any such modifiers.

	<u>Pre-nom adj*</u>	<u>Post-nom adj</u>	<u>Pred adj</u>	<u>Relative cl.</u>
XII-XV	45% (89/199)	2% (3/199)	2% (3/199)	10% (20/199)
XVII-XX	1% (4/452)	1% (4/452)	(1/452)	1% (3/452)

\*includes pre-nominal adjectives counted as determiners (Table 2)

Table 3: Drop in adjectival modification

- (6) e pecado fizo quien le puso en este **grant pesar** (XIV, Zifar, 91)  
tal vez mata una súbita alegría como suele matar un **improviso pesar** (XVII, Persiles y Sigismunda, Capítulo nono del tercer libro)  
esto produce **bastante pesar** más que satisfacción (XX, COREC, AHUM031A.9)
- (7) ca estos anbos han poca alegría en este siglo; desí van a **pesar durable** (XIII, Calila, 293)  
le han ocasionado [...] **pesares graves** (XIX, Bandidos, 369)
- (8) & en cabo com<m>o q<u>i`er q<ue>l **pesar** fuesse **grant** (XIII, GE, fol.248v)  
fue **tanto el pesar** que sintió de ver que os dejaba (XVII, Quijote 2, Capítulo XXIII)
- (9) deziendo sus palabras muy estrañas con grant **pesar que tenia de sus hijos** (XIV, Zifar, 87)  
ca de otra guisa, por el **pesar que estos locos me fizieron**, esso oviera fecho a vós que a ellos (XIV, Lucanor, 201)

A final measure of *pesar*'s loss of noun-hood is a decrease in coordination or juxtaposition with a (non-derived) noun, as in example (10), *más enojos et pesares que plazerres*. Coordination with another noun indicates that *pesar* still conserves lexical meaning. Close to one third, 30%, of Old Spanish *pesar* tokens appear coordinated or juxtaposed with another noun (Table 4). The rate appears identical in 17<sup>th</sup> century texts, but nearly two-thirds (64%) of these are cases of *a despecho y pesar de*, which is itself a collocation in Cervantes's prose (Ex. 11). In 19<sup>th</sup> and 20<sup>th</sup> century data, less than 1%, or virtually none, are coordinated.

	<u>Pesar in coordination – juxtaposition with another noun</u>	
XII-XV	30% (59/199)	
XVII	37% (22/81)	(but 64% (14/22) are <i>a despecho y pesar de</i> )
XIX-XX	1% (2/370)	

Table 4: Drop in coordination – juxtaposition with nouns

- (10) nunca passó día que non oviesse más **enojos et pesares que plazerres**. (XIV, Lucanor, 317)
- (11) comenzó a dar a nuestro don Quijote tantos palos que, **a despecho y pesar de** sus armas, le molió como cibera (XVII, Quijote 1, Capítulo IV)  
no se desesperó de hacer la comedia y de encajar el tal lacayo, **a pesar de** todas las reglas de la poesía **y a despecho del** arte cómico (XVII, Persiles y Sigismunda, Capítulo segundo del tercer libro)

In short, all four measures—plural marking, determination, adjectival modification, and conjoining with nouns—converge on the same story: erstwhile noun *pesar* has shed its nominal trappings. This is a clear case of decategorialization in grammaticization: a noun has been absorbed into a connective.

### 3. From collocation *a + pesar + de* to single unit *a pesar de*

Grammaticization has traditionally been defined as the evolution of lexical into grammatical material (e.g., Hopper & Traugott 1993:xv). Such a definition affords only a partial view, however, since it neglects collocations. Bybee (2003:603) puts forward a characterization of grammaticization that draws attention to the conventionalization of usage patterns involving collocations: “the process by which a frequently used sequence of words or morphemes becomes automated as a single processing unit.” This scholar proposes that the grammaticizing phrase gains autonomy on two planes (Bybee 2003:618). On the one hand, the erstwhile individual components weaken their association with other instances of the same item. In Bybee’s (2003:618) example, as *be going to* reduces to *gonna*, its composite morphemes lose their association with *go*, *to* or *-ing*. On the other hand, the grammaticizing phrase is disassociated from other instances of the construction. Thus, *be going to* loses its association with the more general constructional schema [[movement verb + Progressive] + purpose clause (*to* + infinitive)], as in *I am going/traveling/riding to see the king* (Bybee 2003:603).

Since there is no apparent phonetic reduction as in the case of *gonna*, what kind of diachronic evidence can show increased fusion of *a*, *pesar*, and *de* and greater autonomy of the resulting unit?

#### 3.1 Origins of *a pesar de* in Old Spanish *pesar* constructions

The infinitive nominalization *pesar* originally appeared in a variety of contexts. Most frequently, *pesar* occurred as an object or in verb-object compounds, as in (12) *ovo grant pesar*, or in adverbial phrases with prepositions *con*, *de*, *en*, *por*, as in (13) *con el grant pesar*, and even as a subject or predicate nominal, as in (14) *grandes son los pesares*.

- (12) Quando vio David la villa destroyda , **ovo** grant **pesar** e ploro mucho (XII, Fazienda, 102)  
de que resçebio la dueña muy grant **pesar** (XIV, Zifar, 40)  
que no recibiesse él pena, que él no sentía **pesar** (XV, Celestina, 340)
- (13) Et con el grant **pesar** deste desden dexa se assi morir (XIII, GE, fol.158v)  
Sin seso estava adormida del **pesar** que ove (XV, Celestina, 336)  
Et él estando en este **pesar** et en esta coyta (XIV, Lucanor, 91)  
ca resçibiera mas pesar por el **pesar** que vos ouiesedes (XIV, Zifar, 16)
- (14) Grandes son los **pesares** por tierras de Carrion (XII, Cid, 3697)  
éste es el **pesar** et el cuidado que tengo (XIII, Calila, 285)  
El **pesar** que tengo de tus males te seria satisfacción (XV, Cárcel, 152)

More than half, 56% (112/199), of Old Spanish *pesar* tokens are objects or part of verb-object compounds, especially with *haber*. More than a fourth, 27% (54/199), occur in adverbials, and 10% (20/199) appear in subject or predicate nominal position. This distribution remains fairly steady throughout the Old Spanish period.

The origin of grammaticized *a pesar de* is the [*a + pesar + de + human (sentient being)*] construction, in which *pesar* still means ‘sorrow,’ and the *de + human adnominal* phrase denotes the sentient being who suffers the sorrow. In (15), a man writes a letter offering his soul to the devil, but the letter is reversed thanks to the pleas of a priest and to the sorrow of the devil(s), *a pesar del diablo*, who wept bitterly for the loss of this soul.

- (15) *a + pesar + de + human (sentient being)*  
...fizo al diablo carta de su ánima escrita de su mano, y renegó a Dios poderoso, tomando al diablo por señor por haber una que él mucho amaba, y húbola en esta manera; pero por ruegos de un santo Padre, **a pesar del diablo**, con muchas oraciones le fue su carta visiblemente tornada, llorando los diablos muy agriamente por aquella ánima que perdían (XV, Corbacho, 90)

Associated with this Old Spanish construction is [*a + pesar + possessive pronoun*], illustrated in (16), *a su pesar*. This appears to have been present throughout the history of Spanish, though generally of low incidence (Table 9, ahead).

Both of these Old Spanish constructions are instantiations of a more general [*a + pesar + genitive*] construction schema (17) that means ‘to X’s sorrow or vexation’ (cf. English to X’s chagrin < Fr. *chagrin* ‘sad’).

- (16) *a + pesar + possessive pronoun*  
 si tú fueres allá contra voluntad de mi marido et **a su pesar** (XIII, Calila, 341)

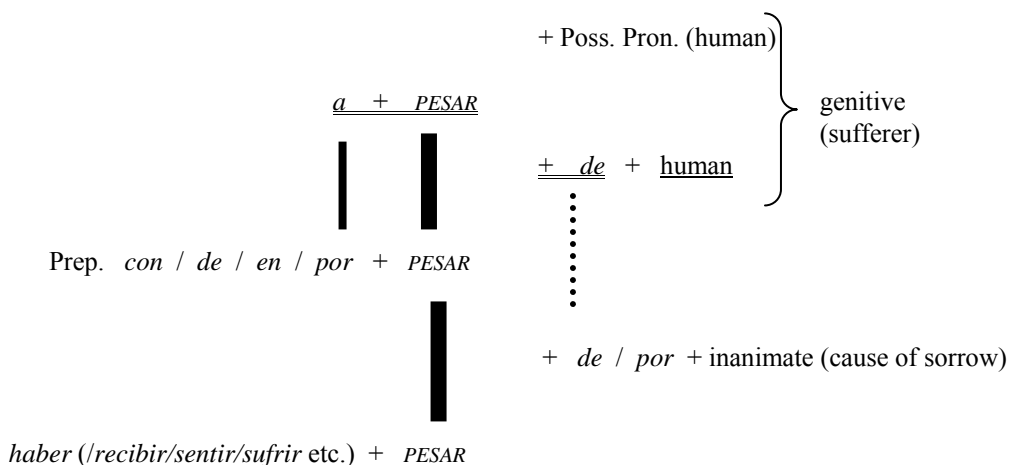
- (17) General constructional schema: [*a + pesar + genitive*]

These and other constructions emerge from distribution patterns involving *pesar* in Old Spanish. We can posit associations, of differing degrees of strength, between *pesar* constructions based on structural and semantic similarities. On the left, the string *a + pesar* itself is associated with a broader [preposition + *pesar*] pattern or construction schema. This is productive in Old Spanish, including prepositions *con*, *de*, *en* and *por* (18). In fact, preceding prepositions other than *a*, which occupies only 5% (10/199), add up to 27% (53/199) of the Old Spanish data. On the right, *pesar + de* is associated with [*pesar + de / por + inanimate*] introducing the cause of the sorrow, often an abstract or deverbal noun (19). Note that the competition between following preposition *de*, with 9% (17/199), and *por*, with 6% (11/199), is balanced.

- (18) e rompieron sos pannos con **pesar** que ovieron e tornaronse a la cibdat (XII, Fazienda, 17)  
 No quiere comer ni beber de **pesar** (XV, Corbacho, 199)

- (19) yo he grand **pesar** desto (XIV, Lucanor, 163)  
 que el rey tenía grant **pesar** por Helbed (XIII, Calila, 298)

Figure 1 depicts these Old Spanish *pesar* constructions and associations. The proposed precursor of grammaticized *a pesar de* is [*a + pesar + de + human*], which is associated with [*a + pesar + possessive*] as part of a more general [*a + pesar + genitive*] construction. On the left, the sequence *a + pesar* is associated with [preposition + *pesar*], and on the right, *pesar + de* is (more loosely) associated with [*pesar + de/por + inanimate*]. Most weakly associated with the precursor of *a pesar de* are instances of *pesar* in verb object position (e.g., *haber pesar*). All these different instances of lexical item *pesar* maintain associations.



Solid line indicates closer association than dotted line; thicker lines indicate associations between instances of lexical item *pesar*.  
 Figure 1: *Pesar* constructions and associations (Old Spanish)

### 3.2 From collocation to unit

The elements of the collocation *a + pesar + de* were initially treated as individual components. Four pieces of evidence can be taken to indicate the compositionality of the collocation: referentiality or “tracking” (Thompson 1997:69), intervening material, juxtaposition with multi-word adverbials, and coordinated adnominal NPs with repeated *de*.

First, there are 17<sup>th</sup> century examples in which anaphoric reference to *pesar* indicates its status as an individual lexical item and hence the compositionality of the entire collocation (in 20, *pesar* and *al* have the same index). It is important that we found no such examples in the 19<sup>th</sup> and 20<sup>th</sup> century data.

- (20) La mujer prisionera, [...], se puso en pie, **a pesar, de sus cadenas y al, de la fuerza** que le hacía para que no se levantase el que con ella venía preso (XVII, Persiles y Sigismunda, Capítulo trece del primer libro)  
Y así habráis de tener paciencia, porque **a vuestro pesar, y al, de vuestro asno**, éste es jaez y no albarda (XVII, Quijote 1, Capítulo XLV)

Other indications of compositionality are cases of material intervening between *a pesar* and *de*, as in (21), *a pesar o quizás a causa de*, and juxtaposition of *a pesar de* with a multi-word adverbial, as in (22), *a pesar de sus bizmas y con dolor de sus costillas*. Juxtaposition with an unambiguously compositional phrase suggests a parallel structure for grammaticizing *a pesar de*, which would still be associated with a more general [preposition + noun + *de*] construction schema (as in, *a causa de, con dolor de*).

- (21) cuyo aspecto, desde el primer instante, le había desagradado de extraño modo, **a pesar o quizás a causa de** que Sabel era un buen pedazo de lozanísima carne (XIX, Pazos, II)
- (22) sentándose en la cama, **a pesar de** sus bizmas y **con dolor de** sus costillas (XVII, Quijote 1, Capítulo XVI)

The strongest evidence for compositionality appears in coordinated adnominal NPs. Repetition of *de*, one *de* for each adnominal NP, shows the relative independence of this component from the other item(s) of the collocation, *a + pesar*. In (23), *de* is repeated with the coordinated NP, *a pesar del recogimiento y de la mansedumbre clericales*, but in (24), one (*a pesar*) *de* suffices, or has scope over, both NPs, *a pesar de los potingues y [NOTHING] las abluciones diarias*. Table 5 indicates a decrease in the repetition of *de*, from an average of 86% in the 17<sup>th</sup> and 18<sup>th</sup> centuries to 60% in 19<sup>th</sup> and 20<sup>th</sup> century data.<sup>2</sup>

- (23) algo de atrevido y varonil en todo el ademán, **a pesar del** recogimiento **y de** la mansedumbre clericales (XIX, Pepita, 316)
- (24) olía a lavanda y espliego, pero por debajo del perfume olía como yo, la fisiología nos igualaba **a pesar de** los potingues **y** las abluciones diarias (XX, Tempestad, 135)

	<u><i>a pesar de X y de Y</i></u>	<u><i>a pesar de X y Y</i></u>
XVII-XVIII	86% (19/23)	14% (4/23)
XIX-XX	60% (18/30)	40% (12/30)

Table 5: Decline of repetition of *de* in coordinated adnominal NPs;  $\chi^2 = 4.298, p = 0.038$

Thus, distribution patterns provide evidence that, over time, erstwhile individual components *a + pesar + de* are fused. This unit of fused elements no longer has a compositional meaning derivable

<sup>2</sup> Results for coordinated adnominal NPs in Table 5 are from supplemental CORDE data. Excluded were tokens in which the second conjunct lacks a determiner (e.g., *a pesar de su extraño nacimiento y sonadas aventuras*, Quijote 1, Cap. VI), since *de* is never repeated in such cases.

from the lexical meaning of *pesar* and the genitive construction, but rather is used as a concessive connective to indicate the relation between two ideas or propositions, as in the typical modern Spanish example (1), *a pesar de que tiene sesenta y cinco metros, parece que es más grande*. As an automated unit (Bybee 2003), associations with other constructions and other instances of *pesar* are severed.

#### 4. Semantic bleaching and syntactic generalization

The automation of *a pesar de* as a single unit (rather than a sequence of preposition, nominal, and adnominal) proceeds in tandem with syntactic generalization of the collocation (cf. Schwenter & Traugott 1995) and semantic bleaching of *pesar* within the emerging unit. As Traugott (2003:638) argues, fixedness in internal structure is accompanied by flexibility in external structure, here realized in the syntactic generalization of *a pesar de* to more classes of nouns, and even to infinitives and finite *que* clauses.

Syntactic generalization is manifested in the element in the erstwhile adnominal *de* phrase. As we have seen, this is originally a human (sentient being) who experiences ‘sorrow, regret.’ Table 6 shows that, in 17<sup>th</sup> century data, *a pesar de* is virtually categorically followed by an NP. In the 19<sup>th</sup> century, NPs still make up two-thirds (66%) of the data, but there is a substantial proportion of infinitives (16%) and *que* clauses (8%). In the 20<sup>th</sup> century, NPs are down to one-third (33%) of all tokens, while *que* clauses alone make up a full third (34%) of the data.

	NP	Infinitive	<i>que</i>	<i>todo/eso</i>
XVII	98% (57/58)	0	0	0
XIX	66% (111/169)	16% (27/169)	8% (13/169)	8% (14/169)
XX	33% (55/167)	10% (17/167)	34% (56/167)	19% (32/167)

Table 6: *a pesar de* + X

How do we get from human adnominals to entire propositions in *que* clauses? Change becomes evident in the 17<sup>th</sup> century, when adnominals are no longer restricted to humans.<sup>3</sup> The extension of the functional range of *a pesar de* occurs as the meaning of *pesar* becomes increasingly abstract. In the 17<sup>th</sup> century examples in (25), notice that *pesar* is not so much ‘sorrow’ or ‘regret,’ as in Old Spanish, but ‘opposition’ by a human to a situation. This is confirmed in the second example by *quisieran estorbarlo*.

- (25) *a pesar de* + HUMAN (opposition)  
 un hombre tan valiente que, **a pesar del comisario y de las guardas**, los soltó a todos (XVII, Quijote 1, Capítulo XXIX)  
 que yo sacara del monesterio, donde, sin duda alguna, debe de estar contra su voluntad, a Leandra, **a pesar de la abadesa y de cuantos quisieran estorbarlo** (XVII, Quijote 1, Capítulo LII)

*Pesar* bleaches even further, going from ‘opposition’ to ‘contrary opinion’. This more abstract meaning is illustrated in (26), where an evaluation, *divinidad de sus ingenios* and *alteza de sus conceptos*, may not meet the approbation of *el circunspecto ignorante que juzga de lo que no sabe*. Though the adnominal is still a human, it is a generic reference to the class of ignorant people.

- (26) *a pesar de* + HUMAN (contrary opinion)  
 muestran la divinidad de sus ingenios y la alteza de sus conceptos, **a despecho y pesar del circunspecto ignorante que juzga de lo que no sabe** (XVII, NE, El licenciado Vidriera)

<sup>3</sup> In first-half 16<sup>th</sup> c. data from CORDE (1500-1511 and 1513-1546, extracted from Libros, Relato extenso novelas y otras formas similares, España; excluding material in verse), adnominals are 100% (27/27) human.

It appears that the first extension from human adnominals is to abstract nouns metonymically related to a human, as in (27). Here *malicia*, *consejos*, and *calumnias* stand for the people holding these. These humans are indicated by an adnominal, for example, *malicia de mis enemigos*, or a possessive, for example, *vuestros consejos*.

- (27) *a pesar de* + METONYMIC HUMAN  
 Tomad mi consejo, y, **a pesar de la malicia de mis enemigos**, casaos con él (XVII, Quijote 2, Capítulo LVI)  
 y quiere seguir su inclinación **a despecho y pesar de vuestros consejos**. (XVII, Persiles y Sigismunda, Capítulo diez y ocho del tercer libro)

Also early is the extension to inanimate entities. These are initially entities that constitute obvious obstacles, for example, in (28), *ungüento*, an obstacle to waking up, or *sombras*, an obstacle to seeing clearly. The examples in (29) illustrate cases that go one step further. Here the entity is evaluated as being potentially in contradiction or incompatible with the idea of the clause. A military uniform may generally be taken to mean militarism or a rosy color good health, but the speaker takes exception to that point of view.

- (28) *a pesar de* + INANIMATE ENTITY (obvious obstacle)  
 Y, en esto, ordenó el cielo que, **a pesar del ungüento**, Carrizales despertase (XVII, NE, El celoso extremeño)  
 y pudo conocer, **a pesar de las sombras** de la capilla, que una de aquellas damas era la Regenta en persona (XIX, Regenta, II)
- (29) *a pesar de* + INANIMATE ENTITY (contradiction/incompatibility)  
**A pesar de mi uniforme**, me desagrada el militarismo (XIX, Perfecta, 208)  
 Yo estoy enferma... sí, señor, **a pesar de estos colores y de esta carne** (XIX, Regenta, XVIII)

A second stage in the syntactic generalization of *a pesar de* is from (metonymic) humans and inanimate entities to abstract nouns or action nominals that encode situations, such as *ligero sueño*, *oscurecimiento*, *matrimonio*, in (30). The incompatibility or contradiction between one's light sleep and another's sneaking out is fairly patent, as is the incongruence between the darkening of colors and their vividness. Perhaps less obvious is the speaker's view of the contradiction between marriage (into one family or crown) and allegiance (to another), in the third example.

- (30) *a pesar de* + ABSTRACT NOUN/ACTION NOMINAL (contradiction/incompatibility)  
 ellas harían con su señora que bajase a escucharle, **a pesar del ligero sueño** de su señor, cuya ligereza no nacía de sus muchos años, sino de sus muchos celos (XVII, NE, El celoso extremeño)  
 la viveza de los colores, apreciable **a pesar del oscurecimiento** producido por la oxidación del barniz original (XX, Tabla, 11)  
 Lee Beatriz de Ostenburga, que, **a pesar de su matrimonio**, por linaje y orgullo de sangre jamás ha dejado de serlo de Borgoña (XX, Tabla, 244)

Semantic bleaching of *pesar* and unit-hood of the collocation are perhaps most manifest when (*a pesar (de)* co-occurs with nouns with like meanings, as in the examples in (31), *sufrimiento*, *disgustos*, and even the same lexical item, *pesares*.

- (31) pero a poco tiempo, y **a pesar de mi tolerancia y sufrimiento**, volvió el citado mi marido a manifestar su anterior conducta (XIX, DLNE 319, 1816)  
 Eran tormentos de la conciencia los que les ofrecía para el caso probable de no salvarse, **a pesar de tantos disgustos** (XIX, Regenta, XII)  
 De modo que, **a pesar de los pesares**, nos tiene usted como siempre, mandados por el infame Barbacana (XIX, Pazos, XVI)



In summary, *a pesar de* evolves from a collocation involving the independent lexical item *pesar* to a concessive connective in which the erstwhile components are fused (*a-pesar-de*). Figure 2 depicts the semantic bleaching of *pesar* within the collocation and the syntactic generalization of the emerging unit. The top of Figure 2 indicates semantic bleaching from referential and concrete to more abstract meaning, proceeding from ‘sorrow-regret’ to ‘obstacle-opposition’ to ‘contradiction-incompatibility.’ The bottom of Figure 2 shows the steps in the syntactic generalization of *a pesar de*, from human adnominals to inanimates, then processes, and finally propositions.

Semantic bleaching of *pesar* within *a pesar de*: SORROW → OBSTACLE → CONTRADICTION

Syntactic generalization of *a pesar de*: HUMAN → METONYM./INANIM. → PROCESS → PROPOSITION

Figure 2: Semantic bleaching and syntactic generalization

Empirical support for the proposed steps in the syntactic generalization of *a pesar de* is provided by the changing distribution of following NPs (Table 7). The relative frequency of human adnominals (Ex. 25-26) drops, from 40% in 17<sup>th</sup> century data to 2% in 19<sup>th</sup>-20<sup>th</sup> century data. The proportion of inanimate entities (combined with abstract nouns metonymically related to a human) (Ex. 27-29) also declines. In contrast, abstract nouns or action nominals such as *oscurecimiento* (Ex. 30) double from 44% to 89%. From processes expressed in action nominals, the erstwhile adnominal goes on to include entire propositions, expressed in infinitives or finite clauses with *que*, as we have seen (Table 6).

	Human	Inanimate/ Metonymic (human)	Abstract
XVII	40% (23/57)	16% (9/57)	44% (25/57)
XIX-XX	2% (4/166)	9% (15/166)	89% (147/166)

Table 7: *a pesar de* + NP

Increasing opacity of internal structure and autonomy, according to Bybee (2003:618), enable new discourse-pragmatic functions. A striking difference between the 19<sup>th</sup> and 20<sup>th</sup> century data is the doubling of the relative frequency of new collocations *a pesar de todo* and *a pesar de eso* (Table 6). Close to one-fifth (19%) of the 20<sup>th</sup> century data are tokens of these expressions. Rather than representing a regress to the earlier nominal-adnominal construction, these are newer discourse marker uses.

*Eso* is vaguely deictic, referring not to an entity but to a clause or series of clauses, as in (32), where *eso* may be referring to *hay kilómetros* and *se controla muy bien*. It is less referential in (33), where *eso* is vaguer: lack of number-gender concordance indicates it does not refer to *explicaciones*. *Todo*, as in (34) *yo soy optimista, a pesar de todo*, does not seem referential at all. Rather, the fixed phrase *a pesar de todo* is more of a stance adverb with an interactional function, akin to English ‘nevertheless.’

*A pesar de eso/todo* is yet one more step away from the concrete meaning of the original collocation. It is important to note that these phrases are not early continuations of the nominal-genitive construction, but a later development in syntactic generalization (with no pre-19<sup>th</sup> century examples in the present corpus). They may well develop into autonomous units in their own right.

- (32) Hay kilómetros y kilómetros de – de – de alambradas. A veces – duro – vas – cómo al llegar se controla muy bien quién llega y tal. Bueno, **a pesar de eso, a pesar de eso**, las autoridades de Estados Unidos calculan que se les cuelan cada año un millón – de inmigrantes ilegales, ¿eh? (XX, COREC, AHUM019A.11)

- (33) -...Porque, por ejemplo, la Alhambra es tan sumamente compleja, que si vas tú sola, pues no te enteras de nada.  
 -No te enteras.  
 -Entonces es...  
 -**A pesar de** las explicaciones de Buendía, de...primero.  
 -**A pesar de eso** ¡je, je! **A pesar**. (XX, Madrid, 378)
- (34) Un comentario así, al margen – Que – yo soy optimista, **a pesar de todo** (XX, COREC, AHUM019A.36)

## 5. Relative frequency of collocations as an impetus of change

It is not surprising that the semantic and morpho-syntactic changes we have chronicled are accompanied by token, or text, frequency increases. The important measure here is the token frequency *not* of *pesar* overall, which indeed fails to show a clear upward trend, but the token frequency of the COLLOCATION. As shown in Table 8, this rises from barely one occurrence per 100,000 words of text in Old Spanish to a normalized frequency of 12 in 17<sup>th</sup>-20<sup>th</sup> century texts, a sharp increase. Though we cannot assume that the discourse contexts that are compatible with the occurrence of *a pesar de* are distributed uniformly throughout the texts sampled, the twelve-fold token frequency increase seems notable enough to be taken to indicate real diachronic change. Indeed, there are evident register or genre differences, with lower token frequencies in the oral 20<sup>th</sup> century corpora than in the 19<sup>th</sup> century novels. There may well also be dialect and social differences (the Madrid *Habla culta* corpus has a normalized frequency of 13 per 100,000 words, while the corresponding Mexico City *Habla culta* has 5 and the Mexico City *Habla popular* shows an even lower 3).

	Word count <sup>1</sup>	<i>pesar</i> – all occurrences		<i>a pesar de</i>	
		N	Frequency <sup>2</sup>	N	Frequency <sup>2</sup>
XII-XV	740,000	199	27	4	.5
XVII-XX	2,815,000	391 <sup>3</sup>	14	342 <sup>3</sup>	12

<sup>1</sup> Rounded down to nearest thousand

<sup>2</sup> Normalized per 100,000 words

<sup>3</sup> Only tokens from texts for which word count available

Table 8: Token frequency: collocation vs. lexical item

We saw earlier (section 3.1) that in Old Spanish *pesar* appeared as a noun in different constructions. However, in 17<sup>th</sup> century data, these contexts of use have shrunk: objects, subjects, and adverbial expressions combined make up a scant 20% (16/81) of *pesar* tokens. And in 19<sup>th</sup> and 20<sup>th</sup> century data, these contexts add up to no more than 3% (11/370). Where did all the *pesar*'s go?

Table 9 shows a spectacular increase in the relative frequency of the string *a pesar de*: from 2% in the Old Spanish data, to 72% in the 17<sup>th</sup> century, 86% in the 19<sup>th</sup> century, and 96% in the 20<sup>th</sup> century.<sup>4</sup> As the string *a pesar de* rises, *pesar* as object, subject, or in an adverbial phrase declines steeply, to the point that it has all but disappeared. In other words, in 20<sup>th</sup> century data, *pesar* occurs virtually always flanked by *a* and *de*.

<sup>4</sup> CORDE data (counting all *pesar*, including infinitive occurrences in verbal periphrases) suggest surges in the increase of *a pesar de* relative frequency between the 15<sup>th</sup> and 16<sup>th</sup> centuries (from 2% (12/523) to 16% (301/1878)) and between the 17<sup>th</sup> and 18<sup>th</sup> centuries (from 22% (432/1921) to 56% (440/783)) and another break between the 19<sup>th</sup> and 20<sup>th</sup> centuries (from 68% (2897/4268) to 85% (3410/4023)). I am grateful to the HLS reviewer who assembled these results.

	<u><i>a pesar de</i></u>	<u><i>a POSS pesar(POSS)</i></u>	<u>Adverbial</u>	<u>Other (OBJ, SUBJ)</u>
XII-XV	2% (4/199)	2% (4/199)	27% (54/199)	69% (137/199)
XVII	72% (58/81)	7% (6/81)	12% (10/81)	9% (7/81)
XIX	86% (169/196)	9% (18/196)	1% (1/196)	4% (8/196)
XX	96% (167/174)	2% (4/174)	1% (1/174)	1% (2/174)

Table 9: Surge in relative frequency of *a pesar de*

Thus, the striking diachronic increase is not the token frequency of lexical item *pesar*, but of the collocation *a pesar de*, with a twelve-fold increase. Change is even more acutely manifested in relative frequency, which swells from 2% to 96%, a remarkable reversal. Relative frequency is also more impervious to genre differences than token frequency.

Why might RELATIVE FREQUENCY be a better measure than token frequency? Relative frequency may be important in promoting the fusion of the items constituting the collocation and at the same time the autonomy of the fused unit from its erstwhile lexical constituent, which is absorbed into the new unit and undergoes morpho-syntactic decategorialization and semantic bleaching. Thus freed up, the new grammatical resource generalizes to more (syntactic) contexts. Where *a pesar de* was once one of many contexts of occurrence of *pesar*, with high relative frequency it takes on a life of its own, or, in Bybee's (2003:618) terms, "becomes automated as a single processing unit."

All three indices that we have tracked, decategorialization (section 2), fusion (section 3.2), and syntactic generalization (section 4), indicate that grammaticization proceeds in tandem with the relative frequency of the collocation. Table 10 shows that co-occurrence of *pesar* with the definite article, repetition of *de* in coordinated adnominal NPs, and the proportion of NPs as the element following *de* all decline, as the relative frequency of *a pesar de* rises.

These data moreover suggest that relative frequency increases may actually precede, rather than follow from, other grammaticization measures. In Old Spanish, shown in the top row of Table 10, the collocation, with negligible frequency, does not yet exist, and *pesar* is a bona fide noun. In the 17<sup>th</sup> century, relative frequency has soared and decategorialization is well advanced. Notice, though, that the relative frequency increase from 2% to 72% is a change of greater magnitude than the four-fold decrease in co-occurring definite articles, which suggests the precedence of relative frequency. Furthermore, unit-hood and especially syntactic generalization are still incipient in the 17<sup>th</sup> century. Even in the 19<sup>th</sup> century, syntactic generalization beyond adnominal NPs seems to lag behind relative frequency. It is attained in the 20<sup>th</sup> century, when the relative frequency of *a pesar de* is overwhelming.

Century	Decategorialization (definite article Table 2)	Unit-hood (repeated <i>de</i> Table 5)	Generalization ( <i>de</i> + NP Table 6)	Relative Frequency (Table 9)
XII-XV	20%	--	--	2%
XVII	5%	73%	98%	72%
XIX	2%	52%	66%	86%
XX	0	30%	33%	96%

Table 10: Grammaticization measures and relative frequency

To summarize, in tracing the evolution of a nominal construction to a (concessive) connective, we empirically established decategorialization of the lexical constituent and then provided evidence that the lexical item is absorbed into a new fused unit. We also saw that loss of internal structure goes hand in hand with increased external flexibility.

The case of *a pesar de* argues for attending to collocations in grammaticization. Furthermore, the evidence points to relative frequency as a better measure of change than token frequency. We conclude that a crucial frequency measure in the grammaticization of collocations is the relative frequency of the sequence of words with respect to the erstwhile individual lexical component. With further empirical study of grammaticization processes, relative frequency may turn out to be not so much a concomitant but an impetus of morpho-syntactic and semantic change.

## Corpus

- Bandidos (1880) = Payno, Manuel. 1945. *Los bandidos de Río Frío*, vol. II, ed. by Antonio Castro Leal. México: Porrúa.
- Calila (1250) = Anonymous. 1987. *Calila e Dimna*, ed. by José María Cacho Bleuca and María Jesús Lacarra. Madrid: Castalia. [86,000 approximate word count]
- Cárcel (1492) = de San Pedro, Diego. 1972. *Cárcel de amor*, ed. by Keith Whinnom. Madrid: Castalia. [25,500]
- Celestina (1499) = de Rojas, Fernando. 1987. *La Celestina*, ed. by Dorothy S. Severin. Madrid: Cátedra. [67,000]
- Cid (ca. 1140) = Menéndez Pidal, Ramón (ed.). 1944. *Cantar de mio Cid: texto, gramática y vocabulario*, vol. 3: texto. Madrid: Espasa Calpe. [30,000]
- Corbacho (1438) = Martínez de Toledo, Alfonso. 1978. *Corbacho*, ed. by Michael Gerli. Madrid: Cátedra. [75,000]
- COREC = Marín, Marcos (dir.) n.d. *Corpus de referencia de la lengua española contemporánea: corpus oral peninsular*. Online: <http://www.lllf.uam.es/~fmarcos/informes/corpus/corpusix.html>. [1,000,000]
- DLNE = Company Company, Concepción. 1994. *Documentos lingüísticos de la Nueva España, altiplano central*. México: UNAM. [260,000]
- Fazienda (12th century) = Almerich, Arcediano de Antiochia. 1965. *La fazienda de ultra mar: biblia romanceada et itinéraire biblique en prose castillane du XIIIe siècle*, ed. by Moshé Lazar. Salamanca: Universidad de Salamanca. [85,500]
- GE (1260-1280) = Alfonso X. 1957. *General estoria: segunda parte*, ed. by Antonio Solalinde, Lloyd A. Kasten, and Victor R.B. Oelschläger. Madrid: CSIC. [263,500]
- Lucanor (1350) = Don Juan Manuel. 1971. *El conde Lucanor o libro de los enxiemplos del conde Lucanor et de Petronio*, ed. by José Manuel Bleuca. Madrid: Castalia. [74,000]
- Madrid (1980) = Esgueva, Manuel, and Margarita Cantarero (eds.) 1981. *El habla de la ciudad de Madrid: materiales para su estudio*. Madrid: CSIC. [140,000]
- México Habla Culta (1970) = Lope Blanch, Juan M. (ed.) 1971. *El habla de la Ciudad de México: materiales para su estudio*. México: UNAM. [167,000]
- México Habla Popular (1976) = Lope Blanch, Juan M. (ed.) 1976. *El habla popular de la Ciudad de México: materiales para su estudio*. México: UNAM. [172,000]
- NE = de Cervantes, Miguel. 2001. *Novelas ejemplares*, ed. by Florencio Sevilla Arroyo. Alicante: Biblioteca Virtual Miguel de Cervantes. Online: <http://www.cervantesvirtual.com>. [181,000]
- Pazos (1886) = Pardo Bazán, Emilia. 1886. *Los pazos de Ulloa*. Alicante: Biblioteca Virtual Miguel de Cervantes. Online: <http://www.cervantesvirtual.com>. [83,500]
- Pepita (1870) = Valera, Juan. 1989. *Pepita Jiménez*, ed. by Leonardo Romero. Madrid: Cátedra. [56,500]
- Perfecta (1876) = Pérez Galdós, Benito. 2001. *Doña Perfecta*. Alicante: Biblioteca Virtual Miguel de Cervantes. Online: [www.cervantesvirtual.com](http://www.cervantesvirtual.com). [65,000]
- Persiles y Sigismunda = de Cervantes, Miguel. 2001. *Los trabajos de Persiles y Sigismunda*, ed. by Florencio Sevilla Arroyo. Alicante: Biblioteca Virtual Miguel de Cervantes. Online: <http://www.cervantesvirtual.com>. [145,000]
- Quijote (1605-1616) = de Cervantes, Miguel. 1996. *Don Quijote de la Mancha*. Miguel de Cervantes: obras completas, ed. by Florencio Sevilla Arroyo and Antonio Rey Hazas, vols. 2 and 4. Madrid: Alianza Editorial-Centro de Estudios Cervantinos. [357,000]
- Regenta (1870-1880) = Alas "Clarín," Leopoldo. 1981. *La regenta*, 2 vols., ed. by Gonzalo Soberano. Madrid: Castalia. [303,500]
- Tabla (1990) = Pérez-Reverte, Arturo. 2001. *La tabla de Flandes*. Barcelona: Plaza & Janés.
- Tempestad (1997) = de Prada, Juan Manuel. 1998. *La tempestad*. México: Planeta.
- Zifar (1320) = Anonymous. 1929. *El libro del cavallero Zifar*, ed. by Charles P. Wagner. Ann Arbor: University of Michigan Press. [33,500]

## References

- Bybee, Joan L. 2003. Mechanisms of change in grammaticization: the role of frequency. *The handbook of historical linguistics*, ed. by Brian D. Joseph and Richard D. Janda, 602-23. Oxford: Blackwell.
- Bybee, Joan L., and Paul J. Hopper (eds.) 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- Hopper, Paul J. 1991. On some principles of grammaticization. *Approaches to Grammaticalization 1*, ed. by Elizabeth Closs Traugott and Bernd Heine, 17-35. Amsterdam: Benjamins.

- Hopper, Paul J., and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge, UK: Cambridge University Press.
- König, Ekkehard. 1985. Where do concessives come from?: on the development of concessive connectives. *Historical semantics, historical word-formation*, ed. by Jacek Fisiak, 263-82. Berlin: Mouton.
- Lapesa, Rafael. 1984. El uso de actualizadores con el infinitivo y la suboración sustantiva en español: diacronía y sentido. *Homenaje a Ana María Barrenechea*, ed. by Ana María Barrenechea and Lía Schwartz Lerner, 65-89. Madrid: Castalia.
- Schwenter, Scott A., and Elizabeth Closs Traugott. 1995. The semantic and pragmatic development of substitutive complex prepositions in English. *Historical pragmatics: pragmatic developments in the history of English*, ed. by Andreas H. Jucker, 244-73. Amsterdam: John Benjamins.
- Thompson, Sandra A. 1997. Discourse motivations for the core-oblique distinction as a language universal. *Directions in functional linguistics*, ed. by Akio Kamio, 59-82. Amsterdam: Benjamins.
- Traugott, Elizabeth Closs. 2003. Constructions in grammaticalization. *The handbook of historical linguistics*, ed. by Brian D. Joseph and Richard D. Janda, 624-47. Oxford: Blackwell.

# Selected Proceedings of the 8th Hispanic Linguistics Symposium

edited by Timothy L. Face and Carol A. Klee

Cascadilla Proceedings Project Somerville, MA 2006

## Copyright information

Selected Proceedings of the 8th Hispanic Linguistics Symposium  
© 2006 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 1-57473-408-3 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.  
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

## Ordering information

Orders for the library binding edition are handled by Cascadilla Press.  
To place an order, go to [www.lingref.com](http://www.lingref.com) or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA  
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: [sales@cascadilla.com](mailto:sales@cascadilla.com)

## Web access and citation information

This entire proceedings can also be viewed on the web at [www.lingref.com](http://www.lingref.com). Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Torres Cacoullós, Rena. 2006. Relative Frequency in the Grammaticization of Collocations: Nominal to Concessive *a pesar de*. In *Selected Proceedings of the 8th Hispanic Linguistics Symposium*, ed. Timothy L. Face and Carol A. Klee, 37-49. Somerville, MA: Cascadilla Proceedings Project.

or:

Torres Cacoullós, Rena. 2006. Relative Frequency in the Grammaticization of Collocations: Nominal to Concessive *a pesar de*. In *Selected Proceedings of the 8th Hispanic Linguistics Symposium*, ed. Timothy L. Face and Carol A. Klee, 37-49. Somerville, MA: Cascadilla Proceedings Project. [www.lingref.com](http://www.lingref.com), document #1253.