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

Rena Torres Cacoullos
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 12th 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 occcasion fue preso Daniel, **a pesar del rey** que lo querie enparar (XII, Fazienda, 179)

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

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.

---

1 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 *pesan* 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 vería en todas partes sería casualidad* (IX, Regenta)), lexicalized nominalizations like *pesan* 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 *pesan* 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 *pesan* 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 *pesan* tokens. Table 1 shows that in 17th, 19th and 20th century data, only three plurals appear in close to 500 tokens. The virtual disappearance of plural marking is one indication of degraded noun-hood.

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>XII-XV</td>
<td>94% (187/199)</td>
<td>6% (12/199)</td>
</tr>
<tr>
<td>XVII-XX</td>
<td>99% (449/452)</td>
<td>1% (3/452)</td>
</tr>
</tbody>
</table>

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 *pesan* 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 19th and 20th century data, only possessives occur with any appreciable frequency, as in (5), *a su pesar*. Cases of *pesan* preceded by any kind of determiner drop to 11% in the 17th, 8% in the 19th, and 3% in the 20th century data.

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

Table 2: Drop in determination

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

(5) Era una mucea 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 *pesan* 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* *pesan*, post-nominal adjective *pesan* *durable*, predicate adjective *el pesar fuese grant*, and relative clause *el pesar que estos locos me fizieron*.
striking contrast, in 17th, 19th and 20th century data, only 12 of nearly 500 tokens (3%) have any such modifiers.

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

* 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 alegria 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 alegria en este siglo; desi van a pesar durable (XIII, Calila, 293)
    le han ocasionado […] pesares graves (XIX, Bandidos, 369)

(8) & en cabo com<m>q<u>i`er q<ue>l pesar fuese 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 fijos (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 plazeres. 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 17th 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 19th and 20th century data, less than 1%, or virtually none, are coordinated.

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

Table 4: Drop in coordination – juxtaposition with nouns

(10) nunca passó día que non oviesse más enojos et pesares que plazeres. (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 grammatization: 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)
esté 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 \([a + \text{pesar} + \text{possessive pronoun}]\), 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 \([a + \text{pesar} + \text{genitive}]\) construction schema (17) that means ‘to X’s sorrow or vexation’ (cf. English to X’s chagrin < Fr. chagrin ‘sad’).

(16) \[ a + \text{pesar} + \text{possessive pronoun} \]

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

(17) General constructional schema: \([a + \text{pesar} + \text{genitive}]\)

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 tenia 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 \([a + \text{pesar} + \text{de} + \text{human}]\), which is associated with \([a + \text{pesar} + \text{possessive}]\) as part of a more general \([a + \text{pesar} + \text{genitive}]\) 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.

[Diagram]

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 \( a + \text{pesar} + \text{de} \) 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 \text{de}.

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

(20) La mujer prisionera, […] se puso en pie, a \text{pesar} de sus cadenas y \text{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 \text{pesar}, y \text{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 \( a \text{pesar} \) and \text{de}, as in (21), \text{a pesar o quizás a causa de}, and juxtaposition of \text{a pesar de} with a multi-word adverbial, as in (22), \text{a pesar de sus bizmas y con dolor de sus costillas}. Juxtaposition with an unambiguously compositional phrase suggests a parallel structure for grammaticizing \text{a pesar de}, which would still be associated with a more general [preposition + noun + \text{de}] construction schema (as in, \text{a causa de}, \text{con dolor de}).

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

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

(23) algo de atrevido y varonil en todo el ademán, a \text{pesar del} recogimiento y \text{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 \text{pesar de} los potingues y las abluciones diarias (XX, Tempestad, 135)

<table>
<thead>
<tr>
<th></th>
<th>a pesar de X y de Y</th>
<th>a pesar de X y Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>XVI-XVIII</td>
<td>86% (19/23)</td>
<td>14% (4/23)</td>
</tr>
<tr>
<td>XIX-XX</td>
<td>60% (18/30)</td>
<td>40% (12/30)</td>
</tr>
</tbody>
</table>

Table 5: Decline of repetition of \text{de} in coordinated adnominal NPs; \( \chi^2 = 4.298, p = 0.038 \)

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

---

2 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., \text{a pesar de su extraño nacimiento y sonadas aventuras}, Quijote 1, Cap. VI), since \text{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 17th century data, *a pesar de* is virtually categorically followed by an NP. In the 19th 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 20th century, NPs are down to one-third (33%) of all tokens, while *que* clauses alone make up a full third (34%) of the data.

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

Table 6: *a pesar de* + X

How do we get from human adnominals to entire propositions in *que* clauses? Change becomes evident in the 17th century, when adnominals are no longer restricted to humans. The extension of the functional range of *a pesar de* occurs as the meaning of *pesar* becomes increasingly abstract. In the 17th century examples in (25), notice that *presar* 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 circunscrito 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* circunscrito ignorante que juzga de lo que no sabe (XVII, NE, El licenciado Vidriera)

---

3 In first-half 16th 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, *vuestras 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* *vuestras 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 de* el *ungüento,* Carrizales despertase (XVII, NE, El celoso extremeno)
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 de* el *ligero sueño* de su señor, cuya ligereza no nacía de sus muchos años, sino de sus muchos celos (XVII, NE, El celoso extremeno)

la viveza de los colores, apreciable *a pesar de* el *oscurecimiento* producido por la oxidación del barniz original (XX, Tabla, 11)

Lee Beatriz de Ostenburgo, 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 17th century data to 2% in 19th-20th 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 *oscurcimiento* (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).

<table>
<thead>
<tr>
<th>Year</th>
<th>Human</th>
<th>Inanimate/Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40% (23/57)</td>
<td>16% (9/57)</td>
</tr>
<tr>
<td>XVII</td>
<td>16% (9/57)</td>
<td>9% (15/166)</td>
</tr>
</tbody>
</table>

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 19th and 20th 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 20th 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-19th century examples in the present corpus). They may well develop into autonomous units in their own right.

(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 17th-20th 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 20th century corpora than in the 19th 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).

![Table 8: Token frequency: collocation vs. lexical item](image)

We saw earlier (section 3.1) that in Old Spanish **pesar** appeared as a noun in different constructions. However, in 17th 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 19th and 20th 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 17th century, 86% in the 19th century, and 96% in the 20th century. 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 20th century data, **pesar** occurs virtually always flanked by **a** and **de**.

---

4 CORDE data (counting all **pesar**, including infinitive occurrences in verbal periphrases) suggest surges in the increase of **a pesar de** relative frequency between the 15th and 16th centuries (from 2% (12/523) to 16% (301/1878)) and between the 17th and 18th centuries (from 22% (432/1921) to 56% (440/783)) and another break between the 19th and 20th centuries (from 68% (2897/4268) to 85% (3410/4023)). I am grateful to the HLS reviewer who assembled these results.
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 17th 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 17th century. Even in the 19th century, syntactic generalization beyond adnominal NPs seems to lag behind relative frequency. It is attained in the 20th century, when the relative frequency of *a pesar de* is overwhelming.

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


Celestina (1499) = de Rojas, Fernando. 1987. La Celestina, ed. by Dorothy S. Severin, Madrid: Cátedra. [67,000]


COREC = Marín, Marcos (dir.) n.d. Corpus de referencia de la lengua española contemporánea: corpus oral peninsular. Online: http://www.lilf.uam.es/~fmarcos/informes/corpus/corpusix.html. [1,000,000]


Fazienda (12th century) = Almerich, Arcediano de Antiochia. 1965. La fazienda de ultra mar: biblia romanceada et itinéraire biblique en prose castillane du XIIe siècle, ed. by Moshé Lazar. Salamanca: Universidad de Salamanca. [85,500]


Lucanor (1350) = Don Juan Manuel. 1971. El conde Lucanor o libro de los enxiemplos del conde Lucanor et de Petronio, ed. by José Manuel Blecua. Madrid: Castalia. [74,000]


References


