Abstract.
The Spanish intransitive motion verbs subir ‘go up’ and bajar ‘go down’ variably appear with the middle marker se. In this study, we examine aspectual (telicity, punctuality) and pragmatic (counter-expectation, speaker involvement) analyses of se-marking. Adopting the variationist method, we operationalize and test hypotheses via quantitative analysis of variation between the middle se-marked and non-se-marked forms in corpora of spoken and written Mexican and Peninsular Spanish. Variability in se-marking is structured, as shown by recurrent patterns reflected in frequencies of co-occurrence. We adduce evidence that factors in speakers’ choice of the middle se-marked form include a focus on the moment of change (operationalized by co-occurring locative preposition), the foregrounding of the situation (clause type), the topicality of the subject (subject expression), and speaker involvement (grammatical person, tense-mood-aspect); that is, a confluence of aspectual and pragmatic factors. Multivariate analysis further reveals that se-marking is most likely to occur in a particular ‘enter-exit a vehicle’ subir-bajar construction. A number of other constructions with clear se-marking preferences also emerge from the data. Rather than generalizations about abstract structures, the empirical results reveal usage-based constructions of differing degrees of lexical specificity and fixedness. A usage-based constructions approach allows us to dispense with a grammar-pragmatics dichotomy.

Keywords: Spanish middle se; Motion verbs; Speaker involvement; Counter-expectation; Transitivity; Variationist method.
Biographical note

Rena Torres Cacoullos is Associate Professor in the Department of Spanish and Portuguese, University of New Mexico. She is the author of *Grammaticization, Synchronic Variation, and Language Contact. Spanish Progressive -ndo Constructions* (Amsterdam: John Benjamins, 2000). She studies language variation and change, especially involving grammaticalization. Recently she has examined, in natural speech data, the relationship between fixed formulas and their associated productive constructions.

Scott A. Schwenter is Associate Professor of Hispanic Linguistics in the Department of Spanish and Portuguese, The Ohio State University. He is the author of *Pragmatics of Conditional Marking: Implicature, Scalarity, and Exclusivity* (New York: Garland, 1999), as well as articles in journals such as *Hispania, Journal of Historical Pragmatics, Journal of Linguistics, Journal of Pragmatics, Language Variation and Change, Lingua, Lingüística Española Actual, Linguistics, Oralia,* and *Studies in Language.* His current research centers on the pragmatics of negation and the comparative analysis of morphosyntactic variation in Spanish and Portuguese.
Pragmatics and constructions: variable middle marking in Spanish *subir*(se) ‘go up’ and *bajar*(se) ‘go down’

Rena Torres Cacoullos and Scott A. Schwenter
University of New Mexico and The Ohio State University

1. Introduction

Morphological marking is traditionally viewed as the prerogative of grammar rather than pragmatics. Spanish *se*, etymologically a third-person reflexive marker, as well as its person/number variants *me* (1sg.), *te* (2sg.), *nos* (1pl.), and *os* (2pl.), variably co-occurs with intransitive verbs of motion, such as *subir* ‘go up’ and *bajar* ‘go down’, as in (1) and (2).

1. Se-marking of motion verbs has been overwhelmingly discussed in terms of aspect and voice, appealing in particular to the notions of telicity, punctuality, transitivity, and middle voice (e.g., Mendikoetxea, 1999:1640 (§26.1.1.3); De Miguel, 1999: 2986 (§46.1.1.2), 2995-2996 (§46.2.3); Clements, 2006:250-256; cf. Hopper and Thompson, 1980:266; Arce-Arenales, Axelrod, and Fox, 1994; Nishida, 1994; Lyons, 1995; Zagona, 1996; MacDonald, 2004). Thus, in (1), ‘going up’ stairs has the middle marker when the situation is perfective and telic (indicated by Preterit morphology and the locative *hasta el último piso* ‘to the top floor’), as in (1a), but not when the situation is imperfective and lacks a built-in endpoint, as in (1b). Similarly, in (2a) *se* may be said to imply the point of origin, here 'the attic', and thus be more bounded than ‘going down’ without *se* in (2b), where the point of origin would seem to be irrelevant to the visibility of the subject’s legs as she descends the stairs (cf. de Molina Redondo, 1974:47-56; de Miguel, 1999:2986-2987; Clements, 2006:254).

(1) a. Y un día, por curiosidad, yo ME subí hasta el último piso (Madrid, 385) ‘and one day, out of curiosity, I SE went up to the last floor’

b. veía unas escaleras como de caracol muy raras y entonces empezaba a subir, cuando llegaba al cuarto piso me caía (COREC, CCON032A) ‘[recounting a dream] I saw like a very strange spiral staircase and then I would start to Ø go up, when I reached the fourth floor I would fall’

(2) a. - Ya les dije que SE bajaran.
- ¿Por qué?
- No es bueno dejarlos solos en la azotea. […]
‘- I told them to SE come down.
- Why?
- It’s not good to leave them alone in the attic. […]’

b. - Quiera Dios que Clotilde baje primero.
- ¿Por qué?

---

1 See Corpora, before References. Examples are reproduced exactly from the corpus cited, except for omitted material indicated by […] and context summarized within [ ]. Ø indicates absence of *se*-marking.
- En la escalera de caracol se le ven las piernas.
(CORDE, Ibargiengoitia, Clotilde en su casa)
‘- God permit that Clotilde Ø comes down first.
- Why?
- On the winding staircase you can see her legs.’

Yet pragmatics has also been incorporated in analyses of Spanish se (e.g., Pérez Vásquez, 2002; Straus, 2003). Particularly intriguing is García’s (1975:158-161) idea that “double mention” via se increases participant involvement, by “concentrate[ing] attention on the participant in focus.” A focusing function is also pivotal in Maldonado’s (1999) proposal that se marks counter-expectation. Developing Kemmer’s (1993) idea that the middle voice is characterized by a low degree of distinguishability among participants, Maldonado (1999) characterizes Spanish se-marked intransitive motion verbs as ‘energetic’ constructions, contrasting with ‘absolute’ constructions (as defined by Langacker [1991:389]). In the energetic construction analysis, the low level of participant distinguishability and the ensuing narrowing of predicational scope yields a focusing function, such that the marker se focuses on the change of state of the experiencer of the motion event; that is, se profiles a particular point in space rather than the entire trajectory of motion (Maldonado, 1999:353-399; cf. Kemmer, 1993:157). For example, Ximena subió la escalera ‘Ximena climbed the stairway’ without se indicates the entire trajectory, while Ximena se subió a la mesa ‘Ximena climbed onto the table’ with se signals a particular point in space, the final point of the trajectory indicated by the a-locative, a la mesa ‘onto the table’ (Maldonado, 1999:368). The posited development of energetic constructions is absolute > focused > unexpected (Maldonado, 1999:390). In the last evolutionary stage, which Maldonado terms pragmaticalization, se comes to be used to mark the conceptualization of the event as counter to normal expectations (Maldonado, 1999:390-394). Grammars of Spanish refer to this as a matiz, i.e. an “extra nuance” (Butt and Benjamin, 2004:377) suggesting untimely or unplanned situations.

We find examples of se-marked subir and bajar in unexpected situations that are undesired, as with the tire that ‘went down (deflated)’ in (3) or the messy pasta that ‘goes up’ in (4), as well as in unexpected situations that are contrary to social conventions, such as going downstairs to dine alone on one’s wedding night in (5) or climbing onto the balcony of the Ministry of Justice in (6).

(3) UNEXPECTED/UNDESIRED
SE le bajó una llanta (MexCult, 125)
‘a tire SE went down (deflated) on him’

(4) -como es muy largo el espagueti se me queda aquí, […] y SE me sube para arriba y me queda todo.
- Y te regañan ¿o no? (COREC, CEDU020D)
‘- [talking about eating spaghetti with chop sticks] since the spaghetti is very long it stays like this, […] and it SE goes up and it’s like all over.
-And do you get scolded, or not?’
UNEXPECTED/CONTRARY TO SOCIAL CONVENTIONS

(5) mi mujer no tenía ganas de cenar, cuando llegamos; pero yo dije, pues, mira pues yo sí, así que (a) ME bajo a cenar solo y (b) bajé a cenar ¡je, je! al comedor, a pesar de recién casado (Madrid, 232)
‘[on their wedding night] my wife didn’t feel like dinner when we arrived [at the hotel]; but I said, well, look I do, so (a) I SE go down to eat alone and (b) I Ø went down to eat, ha ha, in the dining room, in spite of just getting married’

(6) como protesta eh-lo que hemos hecho ha sido subirNOS tres al-al balcón del Ministerio de Justicia (COREC, CCON013F)
‘to protest uh – what we did was to SE climb the three of us onto the balcony of the Ministry of Justice’

Nevertheless, when we examine data from actual speech production, we find variation between middle-marked and non-marked ‘go down’ and ‘go up’ in such counter-expectation contexts. In example (5) above about the wedding night, middle-marked me bajo a cenar ‘I went down to dinner’ (a) is followed immediately by non-marked bajé a cenar ‘I went down to dinner’ (b). Example (7) below, which is the continuation of example (6) about the protest at the Ministry of Justice, illustrates the occurrence of a se-marked subir (a), followed by a non-marked subir (b), in a nearly identical context:

(7) (a) SE han subido arriba. Hemos venido dos disfrazados con un mono, hemos extendido una escalera y los - hemos extendido la escalera y (b) han subido tres arriba. Han desplegado una pancarta y a la media hora o por ahí pues han llegado los guardias jurados y la Guardia Civil y los ha sacado a - a palos prácticamente (COREC, CCON013F)
‘(a) They SE went up. Two of us came dressed up in a smock, we put up a ladder and – we put up the ladder and (b) three Ø went up. They unfurled a sign and after about half an hour the security guards and the Guardia Civil arrived and removed them by force’

Studies of se-marking have conveniently ignored such variation, which belies idealizations of a unique form-function relationship (“one form for one meaning, and one meaning for one form” (Bolinger, 1977:x)). This raises the following question: what is the balance between more grammatical considerations such as aspect and more pragmatic notions such as speaker involvement and counter-expectation in an account of variable se-marking of subir and bajar? And, what method will enable us to answer such a question?

We tackle the question about the relative contribution of grammar and pragmatics to speakers’ choice of se-marked subir and bajar via quantitative analysis of variation between marked and non-marked forms in corpora of naturally-occurring language use. The variationist approach (e.g., Labov, 1969; Poplack and Tagliamonte, 2001:88-95) is preoccupied with the asymmetrical relations between linguistic form and function in actual discourse (Sankoff, 1988a:141). Variation theory confronts this problem of form-
function polyvalence with the hypothesis of ‘neutralization-in-discourse’: while contexts can almost always be found in which different forms have different meanings, there are alternations in which the full accompaniment of meaning distinctions is not pertinent either for the speaker or the interlocutor; neutralization in discourse is the “fundamental discursive mechanism of (nonphonological) variation and change” (Sankoff, 1988a:153-154). Systematic quantitative analysis of repeated alternations, such as (me) bajé a cenar ‘I SE-Ø went down to dinner’ in (5) and (se) han subido arriba ‘they SE-Ø went up’ in (7), enable analysts to discern patterns of co-occurrence with contextual elements. From the variationist viewpoint, the recurrent patterns reflected in frequencies of co-occurrence constitute variable grammatical structure.

We will provide evidence that a focus on the moment of change, or telicity-punctuality, is a factor in speakers’ choice of the se-marked form, as are foregrounding of the situation, topicality of the subject, and speaker involvement. However, we show that the most important factor is the particular construction under consideration: se-marking is most likely to occur in the subir-bajar + a ‘to’/de ‘from’/en ‘in’ + Vehicle construction, that is, when people use subir-bajar to talk about entry into or exit from a vehicle. This result adds to the evidence that grammar includes combinations of prefabricated (Hopper, 1998) or “reusable” (Thompson, 2002:141) fragments. We conclude that the data temper the formalist quest for overarching generalizations about abstract structures such as se-marking. Instead, constructions, of differing degrees of lexical specificity or generality and differing degrees of fixedness or productivity, should be considered basic units of grammar (Bybee, 2006). In contrast to the separation between grammar and usage and between grammar and lexicon assumed in formal approaches to language, a construction-based view of grammar allows us to largely forsake a rigid grammar-pragmatics dichotomy (cf. Kay, 2004).

2. Data

We extracted tokens of bajar(se) and subir(se) from two Mexican and two Peninsular (Spain) Spanish oral corpora totaling over 700,000 words, and from Mexican and Peninsular Spanish plays and novels (see Corpora, before References). Excluded from the quantitative analyses are transitive occurrences (with a co-occurring object), for example, bajar la cabeza ‘lower one’s head’ (N = 193); cases of ‘impersonal’ se (e.g., Butt and Benjamin, 2004:415-417) as in, for example, hay una iglesia también muy bonita [...] en donde se puede subir a... por un elevador hasta la parte alta de la torre ‘there is also a very pretty church [...] where one can go up to...by an elevator up to the high part of the tower’ (MexCult, 199) (N=11); occurrences of the collocation subir y bajar ‘go up and go down’, for example, las callecitas [...] suben y bajan asi ‘the streets [...] go up and down like this’ (MexCult, 48) (N = 95); and cases of cost rising, for example, ¿Que suben las cosas? la gente pide más sueldos ‘Things go up? People ask for raises’ (COREC, CCON004C) (N = 27), which are invariably not marked.² Also

² Only three (of 193) transitive tokens were SE-marked, for example, Porque si no NOS bajábamos el coche ‘Because if we didn’t SE bring (us) down the car’ (Madrid, 349), where se may be said to have a
excluded are nominalizations (puerta de bajar ‘exit door’), reduplications (sube que sube ‘go up and go up’), quoted material (for example, in songs), and tokens with insufficient context (for example, in truncated utterances such as entonces como esperaba llamada digo pues si subo de- si subo y tal ‘so since I was waiting for a call I said well if I go up from–if I go up and stuff’ (COREC, CCON016A). These protocols yielded a dataset of 646 bajar(se) and subir(se) tokens.

Table 1 shows the relative frequency of the se-marked forms in the oral and written corpora. The relative frequency of bajarse and subirse is three times greater in the oral data, at 34% (92/275), than in the written data, at 12% (44/371). The difference in the rate of se-marking between oral and written data underscores the impact genre and register may have in frequency calculations (cf. Biber, 1999). Thus, pragmaticists must be sure not to rely solely on written (or worse, constructed) data for their analyses, since oral, conversational data can be indicative of contexts that do not occur or perhaps do not favor a given variant in written discourse (see note 15).

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Spoken</th>
<th>Written</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bajarse</td>
<td>Subirse</td>
</tr>
<tr>
<td>Mexican</td>
<td>38% (30/79)</td>
<td>41% (28/69)</td>
</tr>
<tr>
<td>Peninsular</td>
<td>27% (19/70)</td>
<td>26% (15/57)</td>
</tr>
<tr>
<td>Totals</td>
<td>33% (49/149)</td>
<td>34% (43/126)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>34% (92/275)</td>
<td>12% (44/371)</td>
</tr>
</tbody>
</table>

Subir(se) and bajar(se) occur with the following four kinds of senses, in the oral and written data: general motion, descending or ascending motion, figurative use (figurative semantic extensions, e.g. rising or falling temperatures), and getting on or off a vehicle. Table 2 displays the distribution of subir(se) and bajar(se) and the rate of se-marking in each category.

3 The difference between the Mexican 39% (58/148) and Peninsular oral data 27% (34/127) in combined bajarse and subirse relative frequency achieves statistical significance at the .05 but not the .01 level (Chi-Square = 4.734019514, p = 0.0296).
Table 2

*Bajar – *subir* uses and *se*-marking frequencies*

<table>
<thead>
<tr>
<th></th>
<th>Proportion of data</th>
<th>Se-marking rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>General motion</td>
<td>31% (203/646)</td>
<td>10% (20/203)</td>
</tr>
<tr>
<td>Descending-ascending</td>
<td>26% (171/646)</td>
<td>20% (34/171)</td>
</tr>
<tr>
<td>Figurative</td>
<td>23% (147/646)</td>
<td>14% (20/147)</td>
</tr>
<tr>
<td>Enter-exit a vehicle</td>
<td>19% (125/646)</td>
<td>50% (62/125)</td>
</tr>
<tr>
<td>[Total]</td>
<td>100% (646/646)</td>
<td>[Average] 21% (136/646)</td>
</tr>
</tbody>
</table>

Most frequent (31% (203/646)) are those we term general motion uses, in which ‘coming’ or ‘going’ may not involve a change in elevation, as with the bus route in (8) or going out to buy something in (9). In these uses, *subir*(se) and *bajar*(se) have quite a general motion meaning, sometimes bleached to auxiliary-like functioning, as indicated by the preposing of the object pronoun clitic *lo* ‘it’ to *bajar* in (9) (cf. Myhill, 1988).

**GENERAL MOTION (GO, COME)**

(8) En las mañanas, *bajaba* a las ocho, y apenas cabían doce personas, y luego *SE* subía, y volvía a *bajar* hasta las cuatro de la tarde, pa llegar aquí a las cinco. Pero era un guayín tirado por animales, por mulas (MexPop, 164)

‘In the mornings it Ø came (down) at eight, and barely twelve people fit, and then it SE went (up), and it Ø came (down) again at four in the afternoon, arriving here at five. But it was a wagon drawn by animals, mules’

(9) *Lo bajo* yo a comprar ahora. (COREC, CCON013D)

‘I’ll Ø go (down) to buy it now’

We also find *subir*(se) and *bajar*(se) in contexts of descending or ascending motion, as when going up a pyramid (10) and going *abajo* ‘down’ (11), in just over one-fourth of the tokens (26% (171/646)).

**DESCENDING/ASCENDING MOTION**

(10) *Suban* a la pirámide. Se ve muy bonito (MexCult, 196)

‘Ø Climb up the pyramid. It looks really pretty’

---

4 General motion includes entering or leaving someone’s house/place/room or a building, for example, *el caso es que he subido y sin quitarme la ropa ni nada es cuando te he llamado* ‘anyway I Ø went up (in) and without taking off my clothes or anything is when I called you’ (COREC, BCON014B)
(11) digo pues ME voy a bajar abajo y si no te localizo, digo, te llamo desde abajo (COREC, CCON016A)
   ‘I said well I will SE come down and if I don’t find you, I said, I’ll call you from down there’

Subir(se) and bajar(se) may mean increasing or decreasing, rising or falling, as for example, with regard to temperature (12) and social or occupational status (13). In such uses, which we term figurative, inanimate subjects are highly frequent (68% (100/147)).

Figurative

(12) ‘ya con... este... medicina y baño de agua fría, SE le bajaba la temperatura ya [...] O, si no, le ponían bolsas de hielo en la cabeza, y hasta que le bajaba. ‘with medicine and a cold bath, his fever SE went down…Or, if not, they would put bags of ice on his head until it Ø went down’ (MexPop, 347)

(13) he ido subiendo poco a poco hasta llegar a jefe (MexCult, 309)
   ‘I’ve been Ø going up little by little until becoming supervisor’

Finally, we grouped together cases of entering or exiting any kind of vehicle, including automobiles, buses, trains, airplanes, and carriages, as in (14) and (15). These cases make up one-fifth (19% (125/646)) of the data.

Exit-enter vehicle

(14) Y bajo del coche descalza (MexPop, 59)
   ‘And I Ø get out of the car barefoot’

(15) ME tengo que subir en ése. En ese camión no me gusta andar (MexPop, 45)
   ‘I have to SE get on that one. I don’t like to go on that bus’

The highest rate of se-marking occurs with entering-exiting vehicles (50% (62/125)) followed by descending-ascending motion (20% (34/171)), while the frequency of se-marking in general motion uses is relatively low (10% (20/203)), as it is for figurative senses (14% (20/147)).

To further examine the data, we adopt the variationist method (e.g., Labov, 1969, 1972a; Sankoff, 1988a, 1988b; Poplack and Tagliamonte, 2001). Our primary analytical tool is Variable-rule analysis, a kind of multivariate analysis which models regularities in large bodies of natural speech data by discovering the set of factor groups which jointly account for the largest amount of variation in a statistically significant way (Sankoff, 1988b; Paolillo, 2002). The contexts of occurrence of subir(se) and bajar(se) are first decomposed into a configuration of independent conditioning factors, whose contribution to speaker choices is then modeled probabilistically in the multivariate analysis. Via these

---

5 Thanks to Sarah Blackwell for pointing out that example 12 illustrates so-called ‘intransitivizing se’ (cf. Whitley, 2002:177-179), usually involving inanimate subjects, which cannot be agents, or animate subjects that are deagentivized by se-marking.
factors, extrapolated from previous analyses of *se* and from our own observations of the data, we operationalize hypotheses about the choice between the *se*-marked and the non-*se*-marked form.

We consider 11 environmental factor groups. Co-occurring preposition is a measure of putative telicity effects; clause type and tense-mood-aspect measure the role of foregrounding; and subject expression may be taken as a measure of topicality, while animacy, grammatical person, and co-occurring datives provide measures of speaker/participant involvement. Also considered are lexical type, construction, and the extralinguistic variables of language variety (dialect) and medium.

3. Results

Table 3 shows a Variable-rule analysis of contextual factors contributing to the choice of *se*-marked *bajar* and *subir*, for 646 tokens (304 *bajar*, 342 *subir*), using the Windows application GOLDVARB X (Sankoff, Tagliamonte, and Smith, 2005). The overall likelihood (corrected mean) for *se*-marking is 0.17, corresponding to an overall rate of 21 percent (in the variable context). Factor groups (independent variables) that contribute a statistically significant effect (*p* = 0.049) are depicted in boldface in the first column: medium, construction, co-occurring locative preposition, grammatical person, and tense. Shown within brackets are those factor groups not selected as statistically significant when considered simultaneously with the other factor groups in the multivariate analysis: clause type, language variety, and verb type. Not included in the multivariate analysis are subject expression and animacy (which interact with grammatical person) and datives (because of the low number of tokens, N = 28). For the significant factor groups, relative magnitude of effect is indicated by the range (cf. Poplack and Tagliamonte, 2001:92-94): of the linguistic factor groups, most important is construction (range = 36), followed by locative preposition (range = 24), subject (grammatical person) (range = 12), and tense (range = 8). The direction of effect is shown by the probability or factor weights assigned to the factors in each group, with values between 0 and 1, such that the closer the factor weight to 1 the more likely the *se*-marked form and, conversely, values below 0.50 favor non-*se*-marked *bajar* or *subir*. We will now discuss these contextual effects, beginning with locative preposition.
Table 3  Factors contributing to the choice of se-marked bajar - subir (significant factors in boldface; [ ] = not selected as significant)

Total N = 646, p = 0.049, Corrected mean .17 (21% se-marked)

<table>
<thead>
<tr>
<th>Medium</th>
<th>Probability</th>
<th>% se</th>
<th>Total N</th>
<th>% data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral</td>
<td>.71</td>
<td>34%</td>
<td>275</td>
<td>43%</td>
</tr>
<tr>
<td>Written</td>
<td>.34</td>
<td>12%</td>
<td>371</td>
<td>57%</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td></td>
<td></td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Construction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enter-exit vehicle</td>
<td>.78</td>
<td>50%</td>
<td>125</td>
<td>19%</td>
</tr>
<tr>
<td>Other uses</td>
<td>.42</td>
<td>14%</td>
<td>521</td>
<td>81%</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td></td>
<td></td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Co-occurring locative preposition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a ‘to’, de ‘from’, en ‘on’</td>
<td>.67</td>
<td>33%</td>
<td>172</td>
<td>28%</td>
</tr>
<tr>
<td>Other or none</td>
<td>.43</td>
<td>16%</td>
<td>439</td>
<td>72%</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td></td>
<td></td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st and 2nd person</td>
<td>.58</td>
<td>31%</td>
<td>204</td>
<td>33%</td>
</tr>
<tr>
<td>3rd and Non-specific</td>
<td>.46</td>
<td>16%</td>
<td>415</td>
<td>67%</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td></td>
<td></td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Tense</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>.55</td>
<td>27%</td>
<td>208</td>
<td>40%</td>
</tr>
<tr>
<td>Non-past</td>
<td>.47</td>
<td>19%</td>
<td>309</td>
<td>60%</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td></td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Clause</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main</td>
<td>[.53]</td>
<td>23%</td>
<td>380</td>
<td>63%</td>
</tr>
<tr>
<td>Subordinate</td>
<td>[.46]</td>
<td>18%</td>
<td>227</td>
<td>37%</td>
</tr>
<tr>
<td>Language variety</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexican</td>
<td>[.52]</td>
<td>23%</td>
<td>365</td>
<td>57%</td>
</tr>
<tr>
<td>Peninsular</td>
<td>[.48]</td>
<td>19%</td>
<td>281</td>
<td>43%</td>
</tr>
<tr>
<td>Verb type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subir ‘go up’</td>
<td>[.52]</td>
<td>22%</td>
<td>342</td>
<td>53%</td>
</tr>
<tr>
<td>bajar ‘go down’</td>
<td>[.47]</td>
<td>20%</td>
<td>304</td>
<td>47%</td>
</tr>
</tbody>
</table>

Log likelihood = -265.586, Chi-square/cell = 1.2350

6 Ns in Total N column (Table 3) do not add up to 646 because of excluded factors (there are too few tokens to include them separately and they cannot be collapsed with other factors): in the locative preposition group, subir-bajar(se) + Infinitive (e.g., bajar a cenar ‘go down to eat’) (N=35); in the subject group, uncodable tokens (subjects of non-finite forms and tokens for which we could not determine specificity) (N=27); in the tense group, non-finite (mostly Infinitive) forms (N=89), Present Perfect or Pluperfect forms (N=39) (29/39 of these in the Peninsular data, with a se rate of 10% (3/29)), and one ambiguous Preterit-Present token; and in the clause group, interrogatives (N= 17), if-clauses (N=6), and uncodable tokens (N=16).
3.1 Focus on change of state - higher transitivity (telicity-punctuality): co-occurring locative preposition

Spanish se-marked motion verbs have been said to focus on the change of state involved in the motion event, a consequence of the low participant distinguishability of the middle voice and the narrowing of predicational scope (Maldonado, 1999:353-398; cf. García, 1975:158-161; Kemmer, 1993:157). In the ‘absolute’ construction, the event develops from beginning to end, whereas in the ‘energetic’ construction with se what is prominent is the pivotal moment of change (Maldonado, 1999:365). Through the postulated focusing function, se contributes dynamicity in the sense of rapidity and, hence, abruptness (Maldonado, 1999:369). For example, subir ‘go up’ and bajarse ‘go down’ are said to be used in “compressed” situations, as in ‘quickly’ in (16) (Maldonado, 1999:371).

(16) Enfila con rapidez a las calles del Olivo y, ya frente a la casa, SE bajan con precipitación (CORDE, Azuela, El tamaño del infierno)
‘He rapidly goes down to the streets of Olivo and, once in front of the mansion, they SE get out quickly’

Nevertheless, the sparseness of co-occurring manner adverbials, all of which occur in the written data, precludes a direct test of the hypothesis that se conveys rapidity and abruptness. The rate of se with a co-occurring ‘fast’ adverbial (for example, apresuradamente ‘hurriedly’, con precipitación ‘hastily’, rápido ‘quickly’) at 6% (1/18) is not higher than the 12% average in the written corpus (Table 1), though there were no cases of se-marking with a ‘slow’ adverbial (0/10) (for example, con sigilo ‘stealthily’, lentamente ‘slowly’, muy despacio ‘very slowly’) in the same corpus.

The cognitive notion of focus on the change of state in Maldonado’s energetic construction analysis is congruent with the notions of telicity and punctuality (Vendler’s (1967) accomplishments and achievements) in aspectual analyses of se as marking relatively high transitivity, in the idea that se implies the source or endpoint of motion (e.g., Clements, 2006:250-256). We operationalize these notions with the co-occurring locative preposition factor group.

The favoring effect of co-occurring a ‘to’, de ‘from’ and en ‘in’, with a Probability weight of .67 (Table 3), provides empirical support for the role of focus on the change of state in se-marking of bajar and subir. With a co-occurring a-locative (17a) the rate of subirse is 33% (31/95), with a de-locative (17b) the rate of bajarse is 28% (9/32), and with an en-locative (17c) the combined rate of subirse and bajarse is 93% (14/15). In contrast, with no co-occurring locative (17d), the overall rate of se-marking drops to 16% (58/360), while with a co-occurring por ‘through, along’ locative (17e), which indicates a path rather than a goal or point of origin, the rate is a scant 3% (1/34).\(^7\)

---

\(^7\) Co-occurring desde ‘from’ or hasta ‘until’, as in the following examples, appear to disfavor se-marking (15% (4/26)) even though they could be argued to indicate telicity, which suggests that the cognitive notion of focus-on-change-of-state may be more appropriate here than the aspectual notion of telicity.
(17)  

a. SE habían subido a la caja, donde estaba la carga del camión (COREC, BNOT004A)  
   ‘They SE had climbed onto the box, where the truck’s cargo was’

b. SE baja de la cama y dice: “Uno, dos, tres, cuatro” (COREC, ALUD023A)  
   ‘She SE gets off of the bed and says: “One, two, three, four”’

c. en canoa sí SE subió, pero nada más de paseo (MexPop, 143)  
   ‘you SE have gotten on a canoe, but only for a ride’

d. tenía una rampa interior para que subieran los coches (MexCult, 199)  
   ‘it had an internal ramp so that the cars Ø could go up’

e. Y luego, hemos bajado por la, por la Costa […] Brava hasta Barcelona. (Madrid, 147)  
   ‘And then, we Ø went down along the Costa Brava until Barcelona’

Note that locative prepositions co-occur with se-marking both in the vehicle construction (section 3.4, below), as in (17c), and outside this construction, as in (17a, 17b). The co-occurring locative preposition effect is in fact independent of construction type: outside the vehicle context, the rate of se-marking is higher with *a*, *de* or *en* than without (22% (24/111) vs. 12% (50/410), Chi-square = 6.369513014, *p* = 0.0116) (with vehicles, the rate of se-marking is essentially the same with a co-occurring *a*, *de*, or *en* (52% (32/61)) as without (47% (30/64)) (Chi-square = 0.389566, *p* = 0.5325)).

3.2. Foregrounding: clause type and tense-mood-aspect

In addition to a focus on the origin or endpoint of movement, a foregrounding function has been attributed to se-marking of motion verbs (Clements, 2006:255-256). Foreground is “the material which supplies the main points of the discourse” (Hopper and Thompson, 1980:280; cf. Labov and Waletzky, 1967; Labov, 1972b). We operationalize the discourse notion of foregrounding by coding for clause type, with the hypothesis that se-marking will be favored in main declarative clauses. Though the direction of effect follows this prediction, with a se-marking rate of 23% in main clauses, compared to 18% in subordinate clauses, clause type as configured does not achieve statistical significance.
(Table 3) (cf. Aaron, 2003:131). Nevertheless, a finer breakdown of clause types indicates that least favorable to se-marking are relative clauses, which show a se-marking rate of only 7% (4/54) (the difference with main clauses (87/380) is significant, Chi-square = 6.843600454, p = 0.0089). If we take relative clauses to encode backgrounded information (Hopper and Thompson, 1980; Goldberg, 2006:130; but see Fox and Thompson, 1990:306), the disfavoring of se-marking in relative clauses such as in (18b) compared to main clauses (18a) supports the hypothesis that se is a (fore)grounding device.

(18) a. Porque después que me dijo el doctor que me bajaran a andar, entóns ME bajé, y este... y a mí se me... se complicó porque me rompieron la vejiga (MexPop, 383)
   ‘Because after the doctor told me they should take me down to walk, well I SE went down, and uhm…and for me…there were complications because they punctured my bladder’

b. a este infierno del que subimos y al que teníamos que caer otra vez...
   (CORDE, Fuentes, La muerte de Artemio Cruz)
   ‘to this hell from which we Ø climbed and to which we had to fall again’

Another measure of foregrounding is co-occurrence with Preterit morphology, given the well-known association between perfective aspect and foregrounding (Hopper, 1979; Hopper and Thompson, 1980:285-286). We hypothesized that se-marking should be favored by the Preterit, as also suggested by García (1975:269, note 43). However, this hypothesis does not hold up: we find that the rate of se-marking is virtually identical for Preterit (past perfective) and Imperfect (past imperfective) verb forms, at 28% (34/120) and 26% (23/88), respectively. Nevertheless, when combined together, these past tense forms favor se-marking with a Probability weight of .55 (Table 3). In contrast, the se-marking rate in Present forms is 16% (37/234). Although the Spanish simple Present covers different aspects (and even tenses), most frequent are stative and habitual uses (Moreno de Alba, 1978:18-41; Torres Cacoullos, 2000:182), which are associated neither with foreground nor with high transitivity; note, too, that the Present is the form of choice for gnomic statements such as los gatos se suben a los árboles ‘cats climb trees’ (Bybee et al., 1994:319). The disfavoring of se by Present compared to Preterit and Imperfect forms may be related to the predictability of habitual and stative situations, whereas counter-expectation is more likely to be overtly indicated on a past (actually realized, specific) situation (see section 3.3.2 below; cf. Aaron and Torres Cacoullos, 2005:619-620).

The highest se rate occurs with Imperatives and Subjunctive commands, at 42% (13/31). This result seems consonant with García’s (1975:159) idea that “double mention [i.e. via se, in addition to the subject] […] concentrates attention on the participant in focus;” double mention via se excludes “other (possibly responsible) parties from attention” which allows the inference of volitionality of the subject (García 1975:160, 182 (note 84)). Imperatives and Subjunctive commands likewise exclude other parties

---

[On how imperfective aspect and atelic Aktionsart reinforce genericity, see Turley (1998).]
from attention and imply volitionality. In contrast with the relatively high frequency of *subir-se-bajar* with Imperatives, it is striking that *se* is wholly absent (0/17) in causatives and other Infinitive constructions where one person requires-permits-helps another to do something. The presence of a requiring-permitting person is incongruous with the moving subject’s volitionality (García, 1975:160). Thus, speakers’ elevated use of *se* with commands, as in (19), and their shunning of *se* in causatives, as in (20), provide empirical support for the pragmatic notion of participant involvement in *se*-marking.

(19) a. *bájaTE*, que Macarena está, tomaros por ahí algo (COREC, ACON006A)
   ‘SE come on (down), Macarena is down there, have a drink or something’
   b. Un borracho-entra en un autobús. Y dice: “¿Cuánto vale el autobús?”
      Dice: “Ochenta pesetas.” Y dice: “Que SE bajen todos; me quedo con él.”
      (COREC, ALUD020A)
      ‘A drunk gets on a bus. And he says: “How much does the bus cost?” He
      says: “Eighty pesetas.” And he says: “Everybody SE get off; I’m buying
      it’

(20) a. Búsquela ahora mismo y *hágale subir* (CORDE, Calvo-Sotelo, La
      muralla)
      ‘Go get him right now and have him Ø come up (in’)
   b. Pero no lo dejaban *bajar* (MexPop, 123)
      ‘But they wouldn’t let him Ø get off [the metro’

3.3 *Topicality and involvement: Subject factor groups*

3.3.1 *Person and animacy*

Evidence for increased speaker involvement with *se* is provided by the results for grammatical person and subject expression. In the multivariate analysis shown in Table 3, first and second person subjects, as in (21a), favor *se*-marking with a Probability weight of .58 and a rate of 31% (64/204); in contrast, third person subjects, as in (21b), show a *se*-rate of 16%.10 Third-person *se*-marking remains lower than first and second person even when we set aside inanimate and non-specific subjects, at 17% (40/229) (Chi-

(21) luego (a) nos *bajamos* a tomar un chato, y (b) él *bajó* con nosotros. (MexCult, 180)
   ‘then (a) we SE came down to have a drink and (b) he Ø came down with us’

With inanimate subjects, as in (22), the rate of *se*-marking is a low 14% (19/137),
compared to 23% (116/495) with human subjects (Chi-square = 11.43102151, p = 0.0007)

---

10 For first person, the *se*-marking rate is 32% (50/154), and for second person, 28% (14/50). The relatively high rate for second person is due at least in part to second person imperatives, which make up about one-third (18/50) of second person tokens, with a *se* rate of 39% (7/18).
0.0156). The few inanimate subjects with se-marking tend to co-occur with a dative with a human referent: 68% (13/19) of se-marked inanimates have a co-occurring first singular me, second singular te, first plural nos, third person singular or plural le(s), as in (23). Thus, se-marking is indeed rare with truly inanimate participants.\footnote{Of the six completely inanimate (without a co-occurring dative) se-marked participants, besides (22b), two are vehicles (see example (8)), one is a plant ‘climbing’ a tree, another is memories climbing up like a plant, and one is wine in the collocation \textit{subirse a la cabeza} ‘go to one’s head’.} Overall, the rate of se with a co-occurring dative is 50% (14/28), which may be taken as a measure of participant affectedness (Maldonado, 1999:394).

\begin{enumerate}
\item \begin{enumerate}
\item como \textit{baja} l’agua ahí los muros... (MexPop, 353)
\begin{flushright}
‘as the water there \(\emptyset\) comes down the walls’
\end{flushright}
\item Como ya SE \textit{había bajado} l’agua, va encontrando su cuerpo (MexPop, 175)
\begin{flushright}
‘As the water (river level) SE \textit{had come down}, he found his corpse’
\end{flushright}
\end{enumerate}
\item \begin{enumerate}
\item Y nosotros, SE nos \textit{han bajado} mucho las ventas desde qu’el mercado sobre ruedas entró (MexPop, 305)
\begin{flushright}
‘And we, sales SE \textit{have gone down for us} a lot since the market on wheels came in’
\end{flushright}
\item creí que allí entre la gente SE me \textit{bajaría} el miedo (CORDE, Rulfo, Pedro Páramo)
\begin{flushright}
‘I thought that there among the people the fear SE \textit{would diminish for me}’
\end{flushright}
\end{enumerate}
\end{enumerate}

3.3.2 Specificity

The observation that speakers eschew se-marked motion verbs in gnomic propositions, for example, \textit{Caen las hojas en otoño} ‘leaves \(\emptyset\) fall in the fall’ as opposed to \textit{Se están cayendo todas las hojas del rosal} ‘all the leaves on the rose bush are SE falling off’, has led to the hypothesis that se-marking does not occur with generic subjects (Clements, 2006:254-255). This particular hypothesis fails to hold up in our analysis. In the present data, non-specific subjects, that is, ones referring to any member of a class of entities (Ashby and Bentivoglio, 1993:69-70), show a 20% (11/56) se rate. While this is lower than for first and second person (31%), it is not lower than the se-marking rate with specific third person human subjects (17%). Overall, se-marking with non-specific subjects follows the general pattern, being highest in the vehicle construction (45% (5/11)), relatively high in contexts of ascending or descending motion (29% (4/14)), and low in general motion and figurative uses (5% (1/19) and 8% (1/12), respectively).\footnote{Examples of a non-specific subject in se-marked \textit{subir-bajar}: in descending-ascending motion, (34b); in general motion, \textit{Pues nosotros tuvimos una clase} […] con D., me figuro, que tenías que \textit{bajar}TE con él al bar, a tomarle el café o el desayuno y hacerle la pelota ‘We had a class […] with D., I think, that you had to SE \textit{go (down)} with him to the bar, to have a coffee or breakfast and kiss up to him’ (Madrid, 418); in figurative use, (29b).} \footnote{Formal definiteness likewise does not appear to correlate with se (cf. Clements, 2006:256-257). Indefinite subject NPs, that is, those with the indefinite article or no determiner, as in (3, 27c), do not show...
The feature of genericity that may be most pertinent here is not the specificity of the subject but the predictability of the predicate. Unexpectedness or undesirability of the situation, as proposed in the energetic construction analysis (cf. Maldonado, 1999:374), rather than subject specificity, seems to distinguish the se-marked from the non-marked form. The consideration of counter-expectation, evidence for which we adduced from the tense effect (see 3.2), is illustrated in the pair of examples in (24). In (24a) a (non-specific) woman getting on a bus is routine, but in (24b) a (non-specific) madwoman getting on a bus to sing or beg deviates from normal expectations (24b).

(24) 

a. me gusta bastante la minifalda, porque es más cómoda para subir a los autobuses (Madrid, 141)
‘I like miniskirts a lot, because they’re easier for Ø getting on the bus’

b. Saludaba […] a mujeres de toda naturaleza, desde la loca que SE sube a cantar o a mendigar en los camiones hasta la refinada (CORDE, Azuela, El tamaño del infierno)
‘He would greet […] women of all kinds, from the madwoman who SE gets on to sing or beg in the bus to the refined one’

3.3.3 Subject expression

Although subject specificity does not appear to influence the choice of subir(se) and bajarse, subject expression, which is variable in Spanish, shows evident correlates with se-marking. Table 4 shows the rate of se-marking by grammatical person and subject expression. Unexpressed subjects (25%) and subject pronouns (41%) show a higher se rate than lexical (full NP) subjects (16%) (the difference in se rate between NP subjects and pronominal or unexpressed subjects combined (27% (88/325)) is significant, Chi-square = 7.833222812, p = 0.0051).

The lower se rate with lexical subjects is due, at least in part, to the skewed distribution of full NPs with respect to medium, grammatical person, and animacy. First, the frequency of full NPs relative to pronominal and unexpressed mentions is higher in the written than the oral corpus (45% (119/267) vs. 24% (56/233), Chi-square = 23.06233, p =0.0000) (cf. Fox, 1987:141). Second, full NPs are overwhelmingly third person—an exception would be examples such as ¿Queréis subir los de la mesa? ‘Do those of you at the table wish to go up?’. Finally, there is a correlation between animacy and anaphoric type (Fox, 1995), as inanimate subjects are more likely than humans to be full NPs (72% (80/111) vs. 25% (94/378), Chi-square = 83.41186, p =0.0000). Since the written medium, third person, and inanimate subjects disfavor se-marking, the disfavoring effect of full NPs is not an independent one. Nevertheless, within the written corpus, the se rate is lowest with full NPs (8% (10/119), compared to 15% (22/148) with pronominal and unexpressed subjects combined, though the difference does not achieve a lower rate of se-marking, at 11% (2/18), than definite subject NPs, at 8% (12/139) (Chi-square = 0.120484631, p = 0.7285); however, the data are sparse.
significance (Chi-square = 2.610773934, p = 0.1061; see note 14) as it is within the third
person class of subjects (14% vs. 21% (30/143), though again the difference is not
statistically significant (Chi-square = 2.480267, p = 0.1153). Besides these skewed
distributions, then, how might we interpret the subject expression pattern?

Unexpressed and pronominal subjects are said to be more ‘topical’ than referents
of lexical subjects. This longstanding intuition in Spanish grammar and more generally
(e.g., the papers in Li, 1976) was verified empirically in the work of Bentivoglio (1983),
who used the Givonian (Givón, 1983) measures of referential distance (the distance in
clauses between a mention of a referent and the immediately preceding mention of that
same referent) and topic persistence (the number of subsequent mentions of a referent in
the following 10 clauses) to reveal that the more ‘reduced’ the form of a Spanish subject
(i.e. null or pronominal), the more topical it was. Thus, the relatively elevated rate of se-
marking with unexpressed and pronominal subjects may be taken to indicate an
association between overtly middle-marked subir(se)-bajarse and topicality (cf. Aaron,
2003:131, who found that se was favored if the subject was coreferential with the subject
of the previous clause(s)).

Table 4.
_Bajar – subir se_-marking by grammatical person and subject expression.

<table>
<thead>
<tr>
<th></th>
<th>First-second person</th>
<th>Third person</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexpressed</td>
<td>28% (38/136)</td>
<td>21% (27/130)</td>
<td>25% (69/279)</td>
</tr>
<tr>
<td>Pronoun</td>
<td>50% (16/32)</td>
<td>23% (3/13)</td>
<td>41% (19/46)</td>
</tr>
<tr>
<td>Full NP (lexical)</td>
<td>(2/3)</td>
<td>14% (21/150)</td>
<td>16% (28/175)</td>
</tr>
<tr>
<td>Invariable</td>
<td>14% (20/146)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>31% (64/204)</td>
<td>16% (56/359)</td>
<td>21% (136/646)</td>
</tr>
</tbody>
</table>

1. Third person column excludes non-specific subjects; vertical Totals column includes non-specific subjects.¹⁴
2. NPs include demonstratives (éste ‘this’, esto ‘that’) and other (non-personal) pronouns, e.g. todos ‘everybody’. The
three first-second person tokens are ¿Queréis subir los de la mesa? ‘Do those of you at the table want to go up’;
algunos nos subimos a un camión ‘some (of us) got on a bus’; lo que hemos hecho ha sido subirnos tres ‘what we did
was to (have) three (of us) climb up’.
3. Invariable: non-expressed subjects of infinitive constructions (e.g., en vez de bajar ‘instead of going up’, al subir
‘upon going up’), causatives (me mandó subir ‘s/he ordered me to go up’), relative clauses (el vapor que sube ‘the
vapor that rises’).

Furthermore, while there is no difference between unexpressed (21%) and
pronominal (23%) third-person subjects, in first and second person, subject pronouns
show a considerably higher se-marking rate, at 50%, than unexpressed subjects, at 28%
(Chi-square = 5.779154, p = 0.0162). First and second person subject pronouns have been

¹⁴ In Table 4, we exclude non-specific subjects given the interaction between referent realization and
specificity in Spanish (cf. Cameron, 1993).
said to add ‘pragmatic weight’ (Davidson, 1996) or mark a ‘referential contrast’ (Detges, 2006). In (25), the subject pronoun yo ‘I’ occurs in a contrastive context, namely, the speaker who would perform the task of changing light bulbs in contrast with her brother who wouldn’t change them. This contrastive context is further marked by the postverbal position of the subject (yo), whose referent (the speaker) is presented in focal contrast with the referent of the subject of the preceding sentence (her brother).

(25) tuvo que poner unas bombillas en un techo muy alto y no se subía. ME tuve que subir yo, lo cual, demuestra lo de la mujer (Madrid, 189)

‘he had to put some light bulbs on a very high ceiling and he wouldn’t SE go up. I had to SE go up, which shows about women’

First and second person subject pronouns may also be taken as indicative of the dynamics of the speaker-hearer dyad. In (26a), nosotros ‘we’ occurs without middle marking following a sí ‘yes’ that responds to the preceding yes-no question proffered by Informant B. However, in (26b), where Informant B appears to have misunderstood the import of Informant A’s response in (a), nosotros co-occurs with middle marking (the 1pl. form nos) following a repeated no, which signals an obviously corrective context (cf. Rudolph, 1996), where Informant A is attempting to clarify her interlocutor’s (mis-)comprehension of the situation in question. In other words, an important function of the se-marking in (26b) is to draw attention to the speaker’s viewpoint in contrast to that of the interlocutor, whose viewpoint is being corrected. Thus, co-occurring yo ‘I’, nosotros ‘we’, usted – tú ‘you’ may be taken as an overt indicator of speaker involvement in García’s (1975:160) sense of concentrating attention on the indicated participant. We conclude that the elevated frequency of se-marking with first and second person subject pronouns provides a measure of how the pragmatic consideration of involvement contributes to speakers’ choice of se-marked subir and bajar.15

(26) Inf. B.—-Sí, está bastante cerca ¿no?
    a. Inf. A.—-Sí, nosotros bajamos andando.
    b. Inf. B.—-O sea bajando (...).
    a. Inf. A.—-No, no, nosotros NOS bajamos andando (Madrid, 348)
      ‘B: Yes, it’s pretty close, right?
      a. A: Yes, we Ø go down on foot.
      B: In other words going down (...).
      b. A: No, no, we SE go down on foot’

In summary, by examining a number of factors related to the subject, we have operationalized topicality and involvement. Higher se rates with unexpressed and

15 Note that the first and second person subject pronoun effect is evident in the oral data only. In the oral data, the se rate for pronominal subjects, at 61% (17/28)) is higher than for unexpressed (33% (49/149); Chi-Square = 7.80596306, p = 0.0052) and lexical subjects (32% (18/56)); in the written data, the average se rate is 15% (20/130) for unexpressed, 11% (2/18) for pronominal, and 8% (10/119) for lexical subjects. Otherwise, the direction of effect (the constraint hierarchies, or ordering of factors) for the factor groups is the same in the oral and written data.
pronominal subjects versus lower se rates with lexical NP subjects suggest that se-marking of subir and bajar is more likely with subjects that are of relatively high topicality. Lower se rates with inanimate and third person human subjects and, conversely, higher rates with first and second person subjects and with datives provide evidence for speaker involvement and participant affectedness in se-marked subir and bajar, as does the higher rate of se with co-occurring first and second person subject pronouns.

3.4 Constructions

Constructions are basically “stored pairings of form and function,” (Goldberg, 2003:219) on a continuum from fixed lexical collocations with or without an open slot, such as subirse a la cabeza ‘go to one’s head’, to schematic patterns such as the ditransitive (double object) construction (he gave her a Coke), Subj [V Obj1 Obj2] (Goldberg, 2003:220); Bybee, 2006:715-716).

The multivariate analysis depicted in Table 3 shows that the linguistic factor group with the greatest magnitude of effect in the choice of se-marked subir or bajar is Construction (range = 36): se-marking is most likely (Probability weight .78) in contexts of exiting or entering a vehicle, be it an automobile, bus, train, or airplane. The average rate of se in this context is 50%. While the rate is higher in the oral (75% (45/60)) than in the written data (26% (17/65)), and in Mexican (60% (50/83)) than in Peninsular Spanish (29% (12/42)), the direction of effect is the same; that is, in both oral and written data and in both Mexican and Peninsular varieties, se-marking is most frequent in the vehicle context.\footnote{In the Mexican oral data, the difference in se-marking rate between the vehicle context (83% (38/46)) and all others (20% (20/102)) is significant (Chi-square = 52.80099, p=0.0000); in the Peninsular oral data, the difference between the vehicle context (50% (7/14)) and others (24% (27/113)) is significant, at the .05 level (Chi-square = 4.330474, p = .0374), but the total number of vehicle tokens (N=14) is low.}

For example, getting out of the vehicle is se-marked, even at the ‘normal’ (habitual, expected) stop, as in (27a), or for non-specific agents pursuing a habit (27b), or for routine trips by non-specific commuters (27c). That is, this vehicle context overrides considerations of counterexpectation.

(27) a. yo ME bajo antes. Normalmente, vamos. (COREC, CCON001)
   ‘I SE get off before. Normally, that is.’

b. la gente SE baja de los coches y compra la droga (COREC, ACON026A)
   ‘people SE get out of the cars and buy drugs’

c. cada vez que un ciudadano de Madrid SE sube en el autobús, en cercanías o en el Metro, el Estado le da 20 pesetas (COREC, APOL023A)
   ‘each time a citizen of Madrid SE gets on the bus, on the local train or on the metro, the State gives them 20 pesetas’

When using subir-bajar to talk about entry into or exit from a vehicle, not only do speakers frequently use se-marking (50% (7/14) of the time in the Peninsular oral data, 83% (38/46) in the Mexican oral data, 26% (17/65) in the combined Peninsular and
Mexican written data), but they also frequently use a prepositional *a* ‘to’, *de* ‘from’, or *en* ‘in, at’ locative phrase (49% (61/125) of the time); in contrast, outside this vehicle context, prepositional *a, de, or en* phrases co-occur with *subir-bajar* only 21% (111/521) of the time). The ‘enter-exit a vehicle’ construction that thus emerges from the data is represented in (28).

(28) **Subir-bajar** enter-exit vehicle construction

\[
\begin{align*}
\text{subir} & \quad + \quad \text{SE} \quad (50\%-83\%) \quad + \quad \text{a} \quad \textit{‘to’ / en \ ‘in’} \quad (49\%) \quad + \quad \textit{‘Vehicle’} \\
\text{bajar} & \quad + \quad \text{de} \quad \textit{‘from’ / en \ ‘in’}
\end{align*}
\]

It is important to point out that the vehicle construction effect is an independent one, not merely a by-product of a co-occurring locative preposition + ‘vehicle’, as already noted above with respect to co-occurring locative prepositions (section 3.1): even without a co-occurring *a, de, or en* phrase, *subir-bajar* when referring to entering or exiting a vehicle shows an average *se-*marking rate of 47% (30/64).

A close look at the data reveals a number of conventionalized word sequences in addition to the vehicle construction. At the more fixed end are idioms with a dictionary (e.g., Real Academia Española) entry: *subirse a la cabeza* ‘go to one’s head; get intoxicated (literally or figuratively)’ (N = 5); and, in the Peninsular data, *subirse a la parra* ‘put on airs; be incensed’ (literally: climb the vine) (N = 3) and *subirsele el pavo* ‘blush’ (literally: the turkey climb up on him/her) (29).\footnote{These three idioms make up virtually all the *se-*marked figurative *subir* cases (section 2). It is interesting that *subirse a la cabeza* is attested in the early *Diccionario de Autoridades* (1726-1739), while a general *subirse* entry does not appear until later editions of the Real Academia Española dictionaries; that is, fixed collocations may emerge at the same time as or even prior to a more general construction.}

(29) **Subir idioms:** SE

a. Esto como no *SE sube a la cabeza* (COREC, CCON019A)
   ‘[Talking about wine] this doesn’t *SE go to one’s head’

b. se han puesto en un plan ya de *subirSE a la parra* todo el mundo
   (COREC, CCON037B)
   ‘everybody wants to *SE put on airs’

c. Me dice que me quiere y *SE le sube el pavo* (CORDE, Gala, Los verdes campos del Edén)
   ‘She says that she loves me and she *SE blushes’

But we also find less fixed, more compositional ‘prefabs’, that is, ‘word sequences that are conventionalized, but predictable in all other ways’ (Bybee, 2006:713). An example is *bajar la regla* ‘get one’s period’ (‘menstruate’; literally: the period come down) (N=4), which occurs without *se* (30). Also, never *se-*marked (N = 0/95) are permutations of the frequent collocation *subir y bajar* ‘go up and down’ (31).
Get one’s period: Ø
estoy tan nerviosa que no me ha bajado (COREC, ACON022C)
‘I’m so worried that it Ø hasn’t come down (I haven’t gotten it) [my period]’

Up and down motion: Ø
suben y bajan los elevadores (CORDE, Azuela, El tamaño del infierno)
‘the elevators Ø go up and down’

Se-marking is also absent in the subir-bajar construction for monetary increases or decreases (N = 0/29) (32).

Cost increase or decrease: Ø
a. ya subieron mucho los precios de... de camión (MexPop, 422)
‘bus prices Ø have gone up a lot’
b. Es que han bajado. Es que hay una reducción de tarifas de hasta el sesenta por ciento (COREC, CCON013G)
‘They Ø have gone down. There’s a rate decrease of up to 70%’

Ascending or descending stairs is overwhelmingly not se-marked (3/66). This specific subir-bajar stairs construction may override general telicity considerations, as when going down an entire staircase (33a) or going up to an endpoint (33b) (cf. Maldonado, 1999:368).

Go up stairs: Ø
a. Total que—que bajé, bueno como pude, las escaleras (COREC, CCON037A)
‘Finally I Ø went down, as well as I could, the stairs’
b. Rodrigo subió los peldaños de azulejo hasta su pieza interior (CORDE, Fuentes, La región más transparente)
‘Rodrigo Ø went up the tiled steps to his interior room’

In contrast, tree climbing is invariably se-marked (9/9), even with a figuratively rising abstract subject, 'memories', when likened to ivy (34a), and even in a gnomic statement with a non-specific subject (uno ‘one’) (34b). Notice that stair and tree climbing are equally telic, in that both imply some endpoint. The difference may be that in today’s world, tree climbing (se-marked) is less routine, more unexpected, and more energy-consuming than stair climbing (non-se-marked).

Go up a tree: SE
a. todos los recuerdos se trastocan, SE suben como una hiedra al tronco macizo de un mezquite (CORDE, Azuela, El tamaño del infierno)
‘all the memories go wild, they SE rise like ivy up the solid trunk of a mesquite’

b. Las higueras yo creo que son los árboles donde más fácil SE sube uno (COREC, CCON034A)
‘Fig tress I think are the trees one most easily SE climbs’
In acknowledging the effects of specific constructions, we still confront patterned variability. That is, variability remains even when taking account of particular constructions. For example, getting in (subir) or out (bajar) of bed is se-marked exactly half the time in the present data (5/10). Nevertheless, the multivariate analysis has shown how the variability is systematically conditioned, that is, “structured heterogeneity” (Labov, 1982:17-18). Let us consider this ‘get in-out of bed’ construction: we hypothesize that the extralinguistic factor of medium contributes to the variation (4/5 oral tokens have se, compared to 1/5 tokens from the written corpus), as may the pragmatic notion of counter-expectation, for example, with jumping up and down on the bed like a maniac or getting out of bed barefoot (35a, b), and co-occurring linguistic elements such as an Imperative (36a) and the preposition de ‘of, from’ (35b, 36c). Of course, there remains some inherent variability, as we have seen, which, however clever the linguistic analysis, cannot be completely factored out.

(35)  

a. entonces SE sube a su cama y ha empezado a dar saltos como una loca (COREC, CEDU029A)  
‘so she SE gets on her bed and she started to jump up and down like a maniac’

b. SE bajan de la cama descalcitos o sucios; lo que sea, tengo que estarlos mirando (MexPop, 430)  
‘they SE get out of bed barefoot or dirty; whatever it is, I have to be watching over them’

(36)  

Dice: “Venga, (a) baja TE y los cuentas” Y (b) baja y dice: “Uno, dos, tres, cuatro - Pues es verdad” No lo has cogido. […] (c) SE baja de la cama y dice: “Uno, dos, tres, cuatro” (COREC, ALUD023A)  
[a joke about a couple in bed counting feet] ‘He says: “Come on, (a) SE get down and count them” And she (b) Ø gets down and says “One, two, three, four—it’s true” No, you’re not getting it. […] (c) She SE gets down from the bed and says “One, two, three, four’’

In summary, the quantitative analysis of variable se-marking of subir-bajar brings to light something previous linguistic analyses have at best dismissed as theoretically uninteresting: a particular construction effect. It would be fair, we think, to characterize a statement such as the following as typical of the sentiment that lexically-particular uses need not concern linguists as much as generalizations about abstract structures:

En cualquier caso, muchas de las formas pronominales con verbos intransitivos aparecen totalmente lexicalizadas, por lo que es muy difícil delimitar qué es exactamente lo que aporta la presencia del pronombre reflexivo… Nuestra clasificación de las oraciones medias no da cuenta de este tipo de oraciones… (Mendikoetxea 1999: 1640).

‘In any case, many of the pronominal forms with intransitive verbs are completely lexicalized, so that it is very difficult to define what exactly the presence of the
reflexive pronoun contributes… Our classification of middle sentences does not account for these kinds of sentences’ (our translation)

The data of actual speech production, however, reveal a number of \textit{subir-bajar} constructions, of differing degrees of fixedness and generality, which show evident tendencies, in some cases categorical, with respect to middle \textit{se}-marking. That is, particular constructions (i.e. pairings of form and experiential meaning) may be conventionalized with or without \textit{se}-marking: we \textit{subir} ‘go up’ stairs without \textit{se} but up trees with \textit{se}. Furthermore, the multivariate analysis (Table 3) shows that the vehicle construction effect is a stronger linguistic constraint on \textit{subir-bajar(se)} variation than co-occurring locative preposition, grammatical subject, or tense.

4. Conclusion

The data analyzed in this paper compel us to recognize the conventionalization of usage patterns in constructions. The strong construction effect, with \textit{se}-marking most favored in the ‘enter-exit a vehicle’ construction, provides evidence that in addition to generalizations based on aspect (telicity-punctuality) and pragmatics (involvement, counter-expectation), grammatical structures may be quite lexically specific. In its view of linguistic structure as emergent from language use, usage-based grammar posits different degrees of conventionalization of discourse patterns, where grammar includes combinations of “prefabs” (e.g., Bybee, 2006:713) or “reusable fragments” (Thompson, 2002:141). The recognition of constructions of differing degrees of lexical specificity contrasts with both the disregard of formal linguistics, in its quest for generalizations about abstract structures divorced from usage, and the perplexity of traditional grammar, in its missing of generalizations.\footnote{Cf. Whitley (2002:184), who states that, “The functions of \textit{se} are varied and complex...”. In recognizing lexical particularities, however, traditional reference grammars are closer to usage than formalist linguistic analyses; for example, Butt and Benjamin (2004:375-376) caution English-speaking students of Spanish that the use of \textit{se} serves “to modify the meaning of the original verb in some often unpredictable way” and advise them to seek “a good dictionary.”}

\textsc{Usage-based} constructions link form and \textsc{experiential} meaning, representing specific lexical items and their contexts of use, including contextual elements usually considered part of pragmatic inferences (cf. Croft, 2001). Cognitive linguists have held that the meaning of words or constructions is encyclopedic: it includes speakers’ knowledge about real world situations or experiences (Croft, 2000:105-107).\footnote{Thus, the view taken here may be more compatible with exemplar than with prototype models of representation (e.g., Bybee and Eddington, 2006:325-327).} Our view of usage-based constructions as form-experiential meaning pairings has affinity with Fillmore's Frame Semantics (Fillmore, 1982, 1985), though we emphasize speakers’ experience of contexts of actual use rather than schematic frames. If “grammar is the cognitive organization of [speakers’] experience with language” (Bybee, 2006:711), a rigid grammar-pragmatics dichotomy is untenable. Crucially, constructions are of differing degrees of fixedness and generality, from the idiom \textit{subirse a la cabeza} ‘go to
one’s head’, to permutations of the subir y bajar ‘go up and down’ collocation, to more open—though with decided se-marking preferences—constructions, for example, for climbing stairs or trees. The encyclopedic and experiential component of meaning (real world experiences) makes constructions appropriate units of analysis for pragmatics.

Overlaying particular constructions are general patterns. It is important that even when tokens of the vehicle construction are excluded from the Variable-rule analysis, the magnitude and direction of effect remain the same, with medium and co-occurring locative preposition selected as significant factor groups.\(^{20}\)

The quantitative analysis of subir-bajar(se) variation reveals a sharp difference between oral and written language in se-marking rates, sending a clear message to pragmaticists on data sources. Rather than using hand-picked examples in order to support our hypotheses, as is typical of much research in pragmatics, we have employed the inductive process of the variationist method to discover the distributional patterns and contextual correlates of se-marking. Based on these results, we have then extracted representative examples from the data in order to exemplify the most important patterns and tendencies in the data. The resulting methodological point should be obvious: multivariate analysis, a technique most typically associated with sociolinguistics, can be profitably employed for pragmatic analyses, especially those needing quantitative verification of qualitative hypotheses.

By operationalizing hypotheses on se-marking of motion verbs and confronting them with the data of actual language use in multivariate analysis, we have confirmed some portions of previous analyses and revised others. We found that the cognitive notion of focus on the moment of change or the aspectual notion of telicity (co-occurring a ‘to’, de ‘from’, en ‘in, on’ locatives) contributes to the choice of se-marked subirse and bajarse, as do the discourse notions of foregrounding and topicality (main vs. relative clauses, unexpressed and pronominal vs. lexical subjects), and the more pragmatic notion of counter-expectation (past tense) and speaker/participant involvement (first and second grammatical person, first and second person subject pronouns, co-occurring datives, imperative morphology). It is these recurrent patterns, shaped by a confluence of aspectual and pragmatic considerations, which constitute the structure of variable se-marking of subir and bajar.

**Corpora**

**Oral**


\(^{20}\) Results of separate Variable-rule analysis excluding all vehicle tokens (N = 521, \(p = .0000\), Corrected mean = .119): Medium, oral .68 vs. written .37 (range = 31); Co-occurring locative preposition, a/de/en .70 vs. other or none .44 (range = 26); Subject, 1\(^{st}\)-2\(^{nd}\) person .57 vs. 3\(^{rd}\) person .47 (n.s.); Tense, past .55 vs. non-past .47 (n.s.); Clause, main .52 vs. subordinate .46 (n.s.); Variety, Peninsular .53 vs. Mexican .47 (n.s.); Verb type, subir .53 vs. bajar .47. The only reversal in direction of effect is in Variety.


**References**


Acknowledgments

Thanks to Joan Bybee, Clancy Clements, Salvador Pons, Sarah Blackwell and two anonymous referees for comments on this paper.