Big data and small dialects:**Transitive *andar* in Central American Spanish**

Shannon Rodríguez and Chad Howe

University of Georgia

Abstract: In this paper, we take a corpus approach to analyzing a low-frequency phenomenon in under-represented varieties of Spanish. A quick comparison of publicly available corpora (e.g., the *Corpus del Español*, Davies, 2016-) reveals a wide disparity between well-represented varieties, such as Mexican Spanish and Peninsular Spanish which represent over a third of total data in the *Corpus del Español*, and less commonly studied varieties, like Salvadoran, Honduran, and Nicaraguan Spanish, which together comprise less than 6%. The data from these lesser-studied dialects, though limited in comparison to other varieties, offer important opportunities to understand otherwise neglected linguistic patterns. To address this gap in the literature, we explore the use the verb *andar* in transitive contexts, as in *Y esta es la razón por la que no **ando** novia* ‘And this is the reason that I **don’t have** a girlfriend’ (example from Twitter; see Hernández, 2002 and Lipski, 1994). These claims are based on a variety of data sources, including the aforementioned *Corpus del Español*. The tokens of *andar* were analyzed using a logistic regression with various grammatical factors serving as predictor—e.g., subject type and animacy of the object. Our analysis reveals that this pattern is unique to Central American Spanish and extending into broader contexts of possession. The over-arching objective with this analysis is to show how tools from corpus linguistics can be leveraged to address linguistic issues in under-sourced varieties.

Keywords: Corpus linguistics, big data, Central American Spanish, lexical change, transitive *andar*, syntactic change

1. Introduction

The analysis of low-frequency phenomena presents a number of interesting challenges for linguists, not the least of which is determining a reliable source of data large enough to provide adequate opportunities for discovering the target structure(s). This challenge expands as one considers other facets of the data coverage, such as dialect diversity, which can complicate an otherwise complex topic. This paper leverages contemporary corpus resources to address an issue that sits at the convergence between low-frequency linguistic phenomena and under-represented dialects in Spanish. Specifically, we provide a corpus-based analysis of *andar* in varieties of Central American Spanish with the objective of demonstrating a lexical shift towards increasingly transitive contexts that is specific to this dialect region. This pattern, referred to by scholars as “transitive *andar*” (Rivas, 1987; Hernández, 2002; Lipksi, 1994, 2008), is illustrated in examples (1) and (2) taken from the *Corpus del Español: Web/Dialects* (CdE, Davies 2016-) and Twitter, respectively.

- (1) *Había que contestarle a la guardia, todos nosotros **andábamos** unas pistolas viejas.*

‘We had to answer to the guard/police, we all carried old guns.’

(Nicaragua, CdE)

- (2) *Y esta es la razón por la que **no ando** novia.*

‘And this is the reason that I don’t have a girlfriend.’

(Twitter, Díaz M. 2014)¹

The lexical shift in *andar* discussed in this analysis can be observed in the fact that both *andábamos* in (1) and *ando* in (2) occur with nominal complements producing a meaning akin to ‘to carry’ or ‘to have’. In fact, the *Diccionario de la Real Académica Española* (DRAE) notes that *andar* can denote the meaning of “[l]levar algo consigo” ‘to carry on one’s person’, specifically in Costa Rica, El Salvador, Honduras, and Nicaragua. In the vast majority of Spanish dialects, the verb *andar* is an intransitive verb of motion—similar to other verbs like *caminar* ‘to walk’, *pasear* ‘to go for a walk’, *moverse* ‘to move oneself’, and *trotar* ‘to jog’—that, at least canonically, occurs with prepositional adjuncts. This use of *andar*, shown in examples (3) and (4), is the basic intransitive pattern with the meaning of ‘to walk, to go’. The prepositional phrases *en grupo* and *en bicicleta* function as adjuncts, making them distinct from the complements in examples (1) and (2) with respect to their relationship with *andar*.

- (3) *O sea que existen Coyotes que **andan** solos, Coyotes que **andan** en grupo*

‘Um, there are coyotes that walk alone, coyotes that travel in groups’

(Colombia, CdE)

- (4) *Mira a otros niños y practica, luego en unos días **anda** en bicicleta.*

‘[She/he] watches the other kids and practices, then in a few days [she/he] is going to ride his bicycle (going riding on his bicycle).’

(Mexico, CdE)

Andar can also occur with an adjectival adjunct with a concomitant meaning of ‘to have been feeling’ as in (5). Here, *andar* most readily collocates with adjectives like *bueno* ‘good’, *malo* ‘bad’, or *alegre* ‘happy’, according to the DRAE. Other uses of *andar* noted in the DRAE include its use as an interjection (6) and as an existential (7).

- (5) *Que cuando uno **anda triste** todo le sale mal.* (Honduras, CdE)

‘When one is sad, everything goes bad for him.’

- (6) *¡Anda! Pero sí, esta canción me encanta.* (Spain, CdE)

‘Wow! But yeah, I love that song.’

- (7) *Andan muchos locos sueltos por la calle.*

‘There are a lot of crazy people loose in the street.’

(DRAE)

Finally, across a number of dialects, *andar* has grammaticalized as a progressive, parallel to *andar* + GERUND as illustrated in examples (8) and (9) (Torres Cacoullos, 2001:444).

Hernández (2002:103) notes that “the verb *andar*, as in *anda buscando algo* ‘he is looking for something’, functions as an auxiliary in progressive and copular constructions, as well as a lexical verb meaning ‘walk, go, and go around.’” This use of *andar* is widespread among dialects of Spanish.

- (8) *Ando buscando a una chica que está embarazada.*

‘I’m looking for a girl that’s pregnant.’

(Dominican Republic, CdE)

- (9) *Desde hace unos meses ando pensando que puede valer la pena intentar explotar esta posibilidad.*

‘It’s been a few months that I’ve been thinking that it might be worth trying to take advantage of that possibility.’

(United States, CdE)

We maintain that the grammaticalization process proposed by Torres Cacoullos (2001) for the *andar* + GERUND construction is distinct from transitive *andar*, and, though it may be possible that these two developments occur in tandem, a broader exploration of the interplay

between the two will be left for a subsequent analysis. With this in mind, our analysis will proceed in Section 2 with a more fine-grained discussion of the transitive properties of *andar* in these varieties focused specifically on determining (i) the range of complements that occur in these contexts and (ii) the meanings that arise with these complements. Section 3 provides a discussion of valency and possession in language change in the context of the observed patterns of lexical shift proposed for transitive *andar*. Section 4 discusses the methods and initial results of the corpus analysis of *andar* using data extracted from the CdE, followed by Section 5 which discusses these results both across four target Central American Spanish varieties (Costa Rica, El Salvador, Honduras, and Nicaragua) and compared to other varieties. A summary of the overall results is given in Section 6 along with conclusions and discussion of the implications for further research.

2. Transitive *andar*

As stated above, *andar* functions as an intransitive verb of movement in all varieties of Spanish, also displaying a number of other constructional behaviors (e.g., use with adjectival adjuncts and gerunds). More to the point, *andar* is an unergative verb whose “single argument...is syntactically equivalent to the subject of a transitive verb” (Sorace, 2000:879) and involves “willed, volitional, controlled acts” (van Gelderen, 2011:108). Thus, in example (3), the two tokens of *andar*, *andan solos* and *andan en grupo*, both have as their subjects the NP *coyotes* which expresses the agentive role.

For this analysis, we distinguish two types of transitive *andar*. The first, seen in (10) and (11), are widespread and available across dialects of Spanish, as are the examples in (3)-(9). This transitive use of *andar* is only used in the sense of ‘traversing a space’ and is similar to the

transitive use of the English verb “to walk”, as in *I **walk the trail** to the top of the mountain.*

What is important about these examples is that the ostensibly nominal complements play same role as other complements headed by the prepositions *por* or *en*, which can be added to these NPs with little change in meaning—i.e. *anda en el camino*. In contrast with examples like (1) and (2), the NPs in examples (10) and (11) function as locative adjuncts.

(10) *Y no tiene metas, solo **anda el camino** despierto.*

‘And he/she doesn’t have goals, he/she just walks his/her path awake.’

(Spain, CdE)

(11) *A estos ladrones, solo les recuerdo que cuando **anden la calle** con su familia y los mire alguno de los que fuimos víctimas de sus estafas.*

‘To those thieves, I only remind them that when they walk the streets with their family, they will see some of us that were victims of their scams.’

(Guatemala, CdE)

Regarding the target phenomenon, Hernández (2002:103) notes “*andar* has also acquired an innovative use that expresses the idea of carrying something on one’s person” (see also Lipski, 1994: 260).” When conveying this meaning, *andar* is used as a transitive (unergative) verb, not entirely dissimilar from verbs such as *llevar* ‘to wear/carry’, *cargar* ‘to carry/load’, *portar* ‘to carry/bear’, *traer* ‘to bring/wear’ etc., all of which can be used to convey the sense of transient possession expressed in (12) and (13). We will refer to these cases of *andar* as marking alienable possession; that is, “those with whom it is possible to in some way sever or terminate the relationship of possession (e.g., through loss, sale or theft)” (Zeshan and Perniss, 2013:7). A similar use of the verb *ta* is observed in Swedish, shown in example (14). Viberg (2010) explains

that *tog* (< *ta* ‘take’) is “both a motion verb and a possession verb” which is under the control of the agent (*hon* ‘she’).

- (12) *Es el maje que, aunque **ande ropa** de marca, galán se le mira, al desnudo, lo maje.*

‘It’s the guy that, even though he may have on (wear) name-brand clothes, one can well see with the naked eye the (his) foolishness.’

(Honduras, CdE)

- (13) *SUBES², ayudará que la gente no **ande dinero** en su bolsa y no sea víctima de robo*

‘SUBES will help so that the people don’t carry money in their bags and aren’t victims of theft.’

(El Salvador, CdE)

- (14) *Hon **tog** med sig en kopp te upp på rummet*

‘She took a cup of tea up to her room.’

Beyond these cases, there is evidence, albeit tentative, that transitive *andar* can be used to indicate inalienable possession as well. In examples (2, repeated below) and (14) and (15) cited from Twitter (Central America), *andar* takes animate NPs *novia* ‘girlfriend’ and *un bebé* ‘a baby’ as a complement with a resultant meaning akin to that of *tener*. Possession in these cases is not necessarily transient and occurs beyond the temporal boundaries of the walking ‘event’. Although a proper understanding of these cases of inalienable possession are important for a broader understanding of the distribution of *andar* in these transitive contexts, we will focus primarily on the uses of *andar* to mark alienable possession.

- (2) *Y esta es la razón por la que **no ando** novia.*

‘And this is the reason that I don’t have a girlfriend.’

(Twitter, Díaz M. 2014)

- (14) *Quiero decirles que ya **ando novia** y me quiere demasiado.*

‘I want to tell you all that I have a girlfriend and she loves me a lot.’

(Twitter, Alexander, 2014)

- (15) *Aunque uds no lo crean me veo igual que @EnitaValle en su foto, con la diferencia que no **ando un bebé** en la panza.*

‘Even though you all don’t believe me that I look the same as @EnitaValle in her photo, with the difference that I don’t have a baby in my belly.’

(Twitter, Turkita, 2016)

The possibility of *andar* to mark inalienable possession notwithstanding, determining what constitutes a transitive use of *andar* is relevant for our current purposes since the meaning that arises is parallel to that expressed with certain prepositional adjuncts. To illustrate this, we present the examples in (16) and (17) all of which express the meaning of ‘to carry’ despite the fact that, in (16a) and (17a), the would-be nominal complement occurs as the object of the preposition *con* ‘with’. We thus use the term “transitive” somewhat loosely when referring to *andar* since we are not claiming that the PPs in (16a) and (17a) and the NPs in (16b) and (17b) have the same syntactic status vis-à-vis *andar*. Moving forward we distinguish these two situations, setting aside examples like (16b) and (17b) as cases of transitive *andar* while treating those with an overt preposition³ as adjuncts.

- (16) a. *¡El que **ande con armas** en la calle sin el permiso, preso!*

‘Whoever carries a gun in the streets without permission, (send him) to jail!’

(El Salvador, CdE)

- b. *¡cómo era posible que alguien que decía predicar un evangelio de paz simplemente **andaba un arma** de fuego para matar!*

‘How was it possible that someone who said they preached the gospel of peace simply carried a fire arm to kill!’

(El Salvador, CdE)

- (17) a. *Yo lo creo, porque el que anda en malos negocios **anda con dinero**.*

‘I believe it, because whoever deals in dirty business has money.’

(Honduras, CdE)

- b. *Disculpe señor, en este momento no **ando dinero**.*

‘Pardon sir, in this moment I don’t have any money’

(Honduras, CdE)

In summary, the argument structure of transitive *andar* is not the canonical structure of VP + PP, such as in examples (3) and (4), where the PP contains a nominal object that, in some cases, is in fact compatible with a possession meaning. Instead, transitive *andar* has the structure of VP + NP, where the absence of the preposition triggers the meaning of alienable possession. Before turning to our corpus analysis, we provide an overview of the issues underlying *andar* as an emergent form of possession in Central American Spanish.

3. *Andar* as lexical change

To begin, we need to provide a clarification regarding the type of structure represented by transitive *andar*. In our analysis, *andar*, like other verbs in Spanish such as *tener* ‘to have’, *llevar* ‘to carry/to wear’, and *cargar* ‘to carry’, is referred to as predicative possession, which has clausal syntax (Zeshan and Perniss, 2013:3). This structure contrasts with attributive possession such as *la casa de Belén* ‘Belén’s house’. Among the different source domains that give rise to possessive constructions, Heine distinguishes eight domains, which he refers to as “event

schemas”, the most important of which, at least for the purposes of the current analysis, are the Action domain (i.e. “what one does”), the Location domain (“where one is”), the Accompaniment domain (i.e. “whom one is with), and the Existence domain (i.e. the being of things around us”) (Heine, 1997, Zeshan and Perniss, 2013:5). According to Zeshan and Perniss, the Action schema “is inherently propositional in nature, containing a transitive verb of taking or seizing” which then “grammaticali[z]es to possessive constructions expressing predicative possession” (2013:6). Our analysis does not make a specific claim regarding the source domain represented by the development of transitive *andar*. Nevertheless, it is certainly feasible that the three aforementioned event schemata could be argued to constitute jointly an approach to modeling the source of the possessive meaning of *andar* in Central American Spanish. We will leave this issue for further investigation.

Moreover, the notion of possession adopted for this analysis requires further comment as well. Lehmann (1998) observes that prototypical possession occurs between a human *possessor* and an inanimate (nonhuman) *possessum*. Thus, our previous example, *la casa de Belén* ‘Belén’s house’, would represent a case of prototypical alienable possession, though it is certainly possible that the possessor is not human (e.g., *la casa del perro* ‘the dog’s house’) or that the *possessum* is animate/human (e.g., *la tía de Belén* ‘Belén’s aunt’). It is also worth noting that inalienable possession implies control over the *possessum* by the possessor in such a way that the *possessum* is “inherently and permanently possessed” (2013:7).⁴ We will not provide a more detailed account of this possible distinction as it relates to transitive *andar* in Central American Spanish, especially in light of the fact that the instances of inalienable possession, inasmuch as they are attested in our data, are relatively uncommon.

One final component of the meaning of *andar*, as an intransitive verb of motion, requires particular comment as we believe it may be useful in understanding how it emerges with this innovative meaning in Central American Spanish. As mentioned above, the verb *andar* is unergative, and, as such, its subject is agentive (see also Perlmutter, 1978 and Levin & Rappaport-Hovav, 1995). We have, however, already noted an important deviation from this characterization with example (5) where *andar* occurs with an adjectival complement—*anda triste* ‘is sad’. This quasi-copular use of *andar* is not unlike Spanish *estar* which gives rise to stage-level interpretations of the type that would be expected with an adjective like *triste*. Similarly, the subject in (5) is not agentive, which is another property of stative verbs like *estar* and *ser* and indeed verbs indicating possession, like *tener* ‘to have’. The point here is that a non-agentive subject, we argue, is a precursor to the development of possessive meaning observed with *andar*. In example (18) we can see that non-animate subjects are in fact attested with *andar* indicating a transient state. It remains to be seen how, if any, the role of the subject may play in the development of the meaning illustrated in (19a) and, importantly, if the subject in (19a) has properties parallel to those of a more canonical verb of possession shown in (19b).

(18) *Se ve que mi teléfono anda mal.*

‘You see that my telephone is not working.’

(Argentina, CdE)

(19) a. *Ando una camisa roja.*

b. *Llevo una camisa roja.*

‘I am wearing a red shirt.’

3.1 *andar* in diachrony

For this brief summary, we will not treat the development of *andar* using the entire analytical apparatus afforded by a grammaticalization approach (Bybee, 2008, Company Company, 2008, Heine and Narrog, 2009, Howe 2018) because the shift in this case is primarily lexical—i.e. related to the selectional restrictions of the predicate. That said, it will be shown later in the analysis that structural changes are indeed observed when we take into account the range of prepositions that co-occur with *andar*. In this way, transitive *andar* constitutes a distinct process of change from that proposed by Torres Cacoulios (2001) for *andar* as a progressive auxiliary. With this in mind, we argue, following Rodríguez (2019), that the case of transitive *andar* in Central America is an example of lexical shift. *Andar* stems from the Romance variant of Latin *ambulāre* (DRAE), which was an intransitive verb (Mahoney 2002) meaning ‘walk, walk about, take a walk’ (Perseus Digital Library). We consider the intransitive uses of *andar*, as exemplified above in (3) and (4), as the prototypical use in Spanish, despite the fact that there are historical instances in which *andar* is ambiguously transitive, as in example (20a).

- (20) a. *pero en la corte no hay bobos: **anda** el dinero en la mano*
 ‘But in the court, there aren’t any fools: money is carried in the hand.’
 (CdE, 17th Century)
- b. *Tenga cuidado con los que **andan** en moto, porque uno nunca sabe.*
 ‘You should be careful with those who ride motorcycles because you never know.’
 (Honduras, CdE)
- c. *No es recomendable ni **andar** una moto tan baja que lo haga a uno encorvarse ni una tan alta que no nos permita apoyar al menos ambas puntas de los pies en el suelo.*

It is not recommended either to **ride/own** a motorcycle so low to the ground that it makes you stoop nor so tall that it does not allow the rider to maintain at least two points of the foot on the ground.’

(CR, CdE)

- d. *Me imagino que tu otro negocio anda bien, porque veo que no sos el único que **anda** moto en Nicaragua.*

‘I imagine that your other business is going well, because I see that you are not the only one who **rides/owns** a motorcycle in Nicaragua.’

(CR, CdE)

This instance of *andar* is, at least relative to broader diachronic changes in Spanish, late and thus may represent a viable candidate for a bridging context in which a novel meaning arises that contrasts with the source meaning (Evans and Wilkins, 1998 and Heine, 2002). The ambiguity in (20a) is both structural and semantic. The (inanimate) subject of the verb *anda* is structurally *el dinero* ‘the money’ with a resultant passive reading that also requires that the subject be understood as an object—i.e. *Se lleva el dinero en la mano (por uno)* ‘Money is carried in the hand (by one)’.⁵ This token represents a critical step in the shift from intransitive to transitive meaning allowing for the wider generalization of transitive uses described in this analysis. In the contemporary data, ambiguities persist despite the fact that the transitive meaning is stable. Consider examples (20b-d). In (20b), the verb *andar* has an animate subject (*los que*) and occurs with a circumstantial adjunct *en moto*, which is common in Spanish for indicating modes of transportation (e.g., *viajo en tren* ‘I travel by train’) and, we argue, does not give rise to the ambiguity discussed in (20a). The infinitival case of *andar* in example (20c), however, does not

have a lexical subject and lacks a preposition to introduce the complement *moto*, factors that we believe gives rise to an ambiguity between the transportation reading (‘ride’) and the possession meaning (‘own’). In the absence of the preposition *en* with *moto*, a similar ambiguity can be observed in example (20d). Again, these types of ambiguities are critical to the type of shift observed with *andar* in Central American Spanish and, as will show, can be tracked through specific structural features, such as the presence vs. absence of prepositions.

3.2 Evidence for lexical change

To reiterate, we propose that transitive *andar* represents a lexical shift in Central American Spanish. In these dialects, transitive *andar* is grammatically productive and generalized to a variety of contexts. For instance, nominal complements can be pronominalized, as in (21), where transitive *andar* is used first with a nominal complement (*possessum*) *el celular* ‘the cell phone’, which is subsequently pronominalized with the clitic *lo* ‘it’ in the next clause.

- (21) *me preguntó que si ella (Yamileth) **andaba** el celular o lo **andaba** yo*
 ‘he asked me if she (Yamileth) **was carrying the cellphone** or if I **was carrying it**’
 (El Salvador, CdE)

By the same token, transitive *andar* can take propositional (i.e. abstract) complements, representing an additional sign of lexical shift. To illustrate this point, examples (22) and (23) demonstrate that *andar* can occur with abstract complements—in this case, temporal complements—in parallel fashion to *llevar*. Compatibility with abstract complements like *año* ‘year’ is, we maintain, an additional indicator of lexical shift.

- (22) *Tía Julia, necesito que me des un consejo con mi ex. **Anduvimos un año** pero nada formal nos peleamos porque yo en realidad quiero algo en serio y él no.*

‘Aunt Julia, I need you to give me advice about my ex. We were together a year, nothing formal, but we fought because I, in reality, want something serious, and he doesn’t.’

(El Salvador, CdE).

(23) *llevo año y medio con mi pareja*

‘I’ve been with my partner for a year and a half’

(Spain, CdE)

One of the mechanisms involved in semantic change, as described by Bybee (2008, 2011), is chunking, referring more specifically to “sequences of units or word strings that are often produced together” (2011: 70). This mechanism, while typically evoked in discussions of structural change, can also play a role in lexical shifts where properties of a surrounding linguistic environment may be absorbed into the meaning of a lexical item. This is the proposal made by Rodríguez (2019), and the one defended here, namely that *andar* forms a strong collocation—a chunk—with *con*. The following examples illustrate what we maintain is the predecessor construction for transitive *andar* in Central American Spanish.

(24) *el que se acompaña con sabios, será sabio, y el que **anda con** recogidos y devotos, será como ellos* (CdE, 16th century)

‘he who accompanies wise men, will be wise, and he who walks with the cloistered and the devout, will be like them’

(25) *y esto todo lo vi yo, y **anda con** mucho dinero* (CdE, 17th century)

‘and all of this I saw, and [he] is walking with [carrying] a lot of many’

(26) *Hice, hermana, resistencia al justicia mayor, que **anda con** orden del rey expresa para prenderme* (CdE, 17th century)

‘I did it, Sister, resistance to the higher justice who was walking with [carrying] an order from the King to arrest me’

Each of these tokens represents a relationship between the underlined NP whose meaning is negotiated, as it were, by the preposition *con*. In example (24), this relationship can only be understood as adjunctival, where *con recogidos y devotos* is understood as a circumstantial modifier. Examples (25) and (26), on the other hand, present two possible interpretations: one is the canonical meaning of *andar* and the other is the meaning innovative meaning of ‘to carry’ or ‘to possess’. These instances of ambiguity are precisely the types of contexts required to give rise to a systematic association directly between the verb *andar* and the NP introduced by *con*, a relationship that, with time, would not require the intervention of the preposition. Bybee (2011: 71) notes that “increases in frequency strengthen the sequential relations within the chunk while weakening the relations of component members to cognates elsewhere.” In this view, a chunk such as *ando con dinero* and *andan con armas*, do not continue to signal the canonical (i.e. intransitive) meaning of *andar*, as in ‘to walk, go, go around’, but rather they represent a new pragmatic meaning, namely ‘to carry on one’s person’, that takes its cue from the semantic properties of its adjunct NP (Roberts & Roussou, 1999). This emergent shift in the structural properties of *andar*, when employed in its innovative form, gives rise to one of the additional mechanisms of semantic shift--i.e. gradual constituent structure and category change. Again, we attenuate the claim of category change since, in the grammaticalization literature, this typically involves a change in lexical class—i.e. verbs to bound morphology (Hopper 1991). We use “category” change in this analysis to refer to change from intransitive to transitive without any implications regarding the status of *andar* as a verb.

Rodríguez (2019) proposes that, following an increase in the frequency of *andar con* around the 16th Century, this collocation began patterning as a phrasal verb. Figure 1 demonstrates the tokens per million of *andar con* taken from the Historical/Genre sub-corpus of the CdE (Davies 2002-) to represent its frequency in all countries until from the 13th to the 20th Centuries, and then from the Web/Dialects sub-corpus (Davies 2016-) to represent the frequency of this structure in all Spanish speaking countries in the 21st Century.

Figure 1: Frequency of *andar con* (“21st SV, ...” refers only to Central American varieties)

Hereafter, reanalysis, most likely occurred as a direct result of chunking, causing a change in constituent structure, from *andar* [*con* NP] to [*andar con*] NP which resulted in a change in the semantic and pragmatic properties of *andar* from ‘walk about, walk’ to ‘carry on one’s person’. Rodríguez maintains that a type of elided *con*, as in (20) above and (27) below, began to appear, about which Rodríguez-Puente (2012) notes that phrasal verbs often develop non-transparent meanings, akin to the notion of loss of compositionality or analyzability (Bybee, 2011). This process occurs with *andar* around the 17th Century and 19th Century where the transitive use first appears and where the frequency of *andar con* starts to stabilize as attested in the data from the *Corpus del Español*.

(27) *El capitán andaba (con) un arma 9 milímetros, oficial, Beretta.*

‘The captain carried a 9-millimeter gun, official, Beretta.’

(*CdE*, 20th Century)

With *andar* (*con*) NP reducing the necessity for *con*, the change in semantic meaning from ‘walk about, walk’ to ‘carry on one’s person’ began to increase. By the 21st Century, *andar con* NP was consistently reduced to *andar* NP, representing the emergent transitive structure, where the innovative, narrowed semantic meaning of ‘carry on one’s person’ is intended. In contemporary Central American Spanish, this meaning has expanded to include alienable possession, akin to *tener* ‘to have’.

To review, this diachronic sketch is intended to provide a general starting point for the corpus analysis provided in the forthcoming section, and there are two primary conclusions to draw. First, transitive *andar* is a lexical change in contemporary Spanish, albeit one with a limited dialectal distribution as will be demonstrated shortly. Second, the source structures for transitive *andar*—intransitive *andar* and the collocation *andar con*—continue to be used widely across Spanish dialects. As we will see, it is only in select dialects—namely Central American varieties—that we observe competition between *andar con* and its modern predecessor transitive *andar*.

4. Transitive *andar*: A corpus approach

Working with low-frequency phenomena from a functionalist perspective requires access to a large and reliable data set. Added to this is the complication of finding adequate coverage of under-represented varieties, particularly those with limited institutional prestige. Fortunately, recent additions to the collection of Spanish corpora have made it possible to expand our inquiries beyond those that are possible with smaller, limited corpora. This increased

functionality greatly facilitates our ability to look for otherwise difficult to find patterns among varieties of Spanish that typically receive less attention in the scholarly literature. Our analysis makes use of the 2 billion-word *Corpus del Español* (Davies, 2016-) and attempts to address the following research questions:

1. What is the dialectal distribution of transitive *andar*?
2. Which linguistic variables are the best predictors for transitive *andar*--specifically regarding nouns that convey transient possession, where alienable possession refers to a non-permanent status?
3. How does transitive *andar* compare to other same sense verbs, like *llevar*?

In this section, we take a quantitative, corpus-based approach to the study of transitive *andar* in seeking to discover patterns of co-occurring forms within particular variable contexts (Torres Cacoullos, 2011). The results are divided into two different sections, one in which we compare the distribution of transitive *andar* across a selection of Spanish dialects and another that compares transitive *andar* to semantically similar verbs, specifically *llevar*. The results of these two analyses explain distributional facts for transitive *andar* and help to demonstrate the utility of a ‘Big Data’ approach to dialect variation and, from a diachronic perspective, lexical change.

Before turning to our analysis, we offer a few comments here regarding the coverage of under-studied varieties of Spanish. Among the options available for cross-dialectal analysis of lexical variation, there are only a few that provide wide coverage across all Spanish-speaking countries, including the *Proyecto para el Estudio Sociolingüístico del Español de España y América* (PRESEEA, Moreno Fernández, 1996, 2005, Moreno Fernández and Cestero, 2020) the *Corpus del Español del Siglo XXI* (CORPES XXI), the *Corpus del Español* (Davies, 2016-), and Sketch Engine’s esTenTen18 (Kilgarriff *et al.*, 2014). Each of these sources provide wide

coverage of different varieties of Spanish, albeit at a remarkably larger scale with the latter two (2 billion and 17.6 billion, respectively). In Table 2 below, we provide summary data from three different corpora, *CORPES XXI*, *Corpus del Español*, and esTenTen, comparing three high-resource dialects--Spain, Mexico, and Argentina--and four target varieties in Central America--Costa Rica, El Salvador, Honduras, and Nicaragua. As expected, the target varieties comprise only a small portion of the overall tokens in the three corpora: 5.2% for *CORPES XXI*, 7.2% for the *Corpus del Español*, and >1% for esTenTen. Despite being the largest of these corpora by far, the esTenTen dataset provides relatively minimal access to these varieties. Taken together, the overall number of tokens and the percentage of coverage available through the *Corpus del Español* makes this an ideal tool for analyzing lexical variation.

	CORPES XXI	Corpus del Español	esTenTen
Spain	133.1 (34.7%)	459.3 (21.9%)	3.6 <i>billion</i> (17.5 %)
Mexico	41.2 (10.7%)	260.6 (12.4%)	1.5 <i>billion</i> (7.2%)
Argentina	34.6 (9%)	182.7 (8.7%)	1.8 <i>billion</i> (8.8%)
Costa Rica	5.1 (1.3%)	31.7 (1.5%)	94.7 (.46%)
El Salvador	5.1 (1.3%)	39.1 (1.9%)	18.3 (.09%)
Honduras	5 (1.3%)	38.9 (1.9%)	13.4 (.07%)
Nicaragua	5 (1.3%)	34.6 (1.9%)	40.3 (.19%)

Table 2: Comparison of three Spanish corpora by number of tokens (words) and percentage relative to overall size

4.1 Methodology

To quantify the dialectal variation between transitive and intransitive *andar*, we extracted tokens from the *Corpus del Español* Web/Dialects sub-corpus (Davies 2016-). Specifically, we isolated collocates of [andar] + (*dinero* ‘money’, *plata* ‘silver/money’, *efectivo* ‘cash’, *dinero en efectivo* ‘cash’, *en efectivo* ‘cash’, *moneda(s)* ‘money/coins’, *arma(s)* ‘weapon(s)/gun(s)’, *pistola(s)* ‘gun(s)’, *ropa* ‘clothes’, *zapato(s)* ‘shoes’, *pantalón(es)* ‘pants’, *camisa(s)* ‘shirt’, *short(s)* ‘shorts’, *tenis* ‘sneakers’, *chanclas* ‘sandals’) in different subcorpora: Central America (Honduras, El Salvador, Costa Rica, Panama, Nicaragua, and Guatemala), Spain, Argentina, Chile, Mexico, and the US. These particular nouns were chosen for their compatibility with alienable possession nature and are based on common nouns found with other same-sense transitive verbs (Hernández 2002). We began with 1104 tokens, and, after removing duplicates,

cases of *andar* + GERUND, *andar* + adjective, and *andar* + *sin* ‘without’ (a negative preposition that makes the verb transitively ambiguous), we were left with 455 useable data points.

These tokens were then coded for the following variables: Country, Collocating Noun, Transitivity (transitive ‘no preposition’ or non-transitive ‘preposition intervening between verb and noun’), Verb Form (tense), Habituality (habitual or not), Preposition (if present, which preposition was used), Determiner (if present before noun, which determiner was used), Adjective (if present, which adjective modified the noun), and Polarity (negative or positive). Transitivity represents the response variable and was, for the purposes of this analysis, coded determined based on the presence or absence of the preposition of an intervening preposition between *andar* and one of the following target nouns. In this way, *ando dinero* was coded as transitive while *ando con dinero* was not. The data was then subjected to a generalized linear mixed-effects model (GLMM) in RStudio using the lme4 package (Bates *et al.*, 2015). The GLMM is a model statistically robust enough to handle non-normal, categorical data and also has predictive power, in that log-likelihoods can be extracted from them for different variable coefficients. This allowed us to treat Transitivity as the response variable for the fixed effects, and Noun as the random effect, testing which of the other variables were best suited as the fixed effect operators.

4.2 Results for analysis of dialectal variation

An overview of the data, as seen in Table 3, reveals that there is a high degree of variation observed with the transitive and non-transitive use of *andar* in Costa Rica, El Salvador, Honduras, and Nicaragua, but not in samples from Guatemala and Panama. There is also little to no variation in the other countries represented in this sample.

Country	Non-transitive	Transitive
Argentina	35 (100%)	0 (0%)
Chile	70 (96%)	3 (4%)
Costa Rica	32 (52%)	29 (48%)
El Salvador	15 (37 %)	25 (63%)
Spain	31 (94%)	2 (6%)
Guatemala	15 (79%)	4 (21%)
Honduras	23 (47%)	26 (53%)
Mexico	48 (88%)	6 (12%)
Nicaragua	19 (44%)	24 (66%)
Panama	10 (100%)	0 (0%)
United States	36 (95%)	2 (5%)
TOTAL	334 (73%)	121 (27%)

Table 3: Results of raw tokens per country

The best fit model included Transitivity as the response variable, Noun as the random variable, and Country and Polarity as predictor variables. These variables, along with the rest of the variables in the data, proved to be statistically insignificant ($p=.9998$ in most cases). Figure 1 demonstrates the significance of Transitivity by Country where the results of a Chi-squared test shows no significant difference between the non-transitive and transitive forms in Nicaragua, Honduras, El Salvador, and Costa Rica (Nicaragua: X-squared = 0.5814, $df = 1$, $p = 0.4458$; Honduras: X-squared = 0.18367, $df = 1$, $p = 0.6682$; El Salvador: X-squared = 2.5, $df = 1$, $p = 0.1138$; Costa Rica: X-squared = 0.14754, $df = 1$, $p = 0.7009$). This corroborates the competing variation

between transitive *andar* and non-transitive *andar*. Figure 2 shows the results of Transitivity by Polarity, where the affirmative tokens favor the transitive form.

Figure 2: Transitivity by country

Figure 2: Transitivity by polarity

Taking a closer look at the four countries where transitive *andar* competes with its canonical (i.e. intransitive) counterpart, we calculated the log-odds that a speaker from each of these countries would use transitive *andar*, based on the best predictor variables given by the GLMM, namely Country (El Salvador, Nicaragua, Costa Rica, and Honduras) and Polarity (affirmative). The results were as follows, where the percentage represents the probability transitive *andar* would be used. Observe that the probability is greater than 50% in each these subsamples of the data that *andar* will be used with a transitive sense, as operationalized by the lack of an intervening

preposition. Also important in examples (28)-(31) is the meaning of alienable possession that is expressed.

(28) **El Salvador** = 71%

*Tenía su faldita y blúmer abajo, a la altura de las rodillas. Él **andaba un short** y el zipper lo tenía abajo.*

‘She had her little dress and bloomers underneath, to the length of her knees. He was wearing shorts and he had the zipper down.’

(29) **Nicaragua** = 63%

*...caminar con garbo de bailarín como si **anduviesen la ropa** más holgada del mundo.*

‘to walk with the grace of a ballet dancer as if they wore the most loose-fitting clothes in the world.’

(30) **Costa Rica** = 57%

*...correr el riesgo de chocar así con un infeliz de estos. ¿Y si **anda un arma**? ¿Y si no le importa y sigue su marcha?*

‘...to run the risk of running into one of those wretched people. And if he has a gun? And if he doesn’t care and continues his march?’

(31) **Honduras** = 57%

*Es conveniente **andar siempre dinero en efectivo**, especialmente billetes de baja denominación, ya sea en lempiras...*

‘It’s convenient to always carry cash, especially bills of smaller amounts, even in Lempiras.’

Regarding preposition variation, we considered whether the tokens from these countries displayed variation among preposition usage, which may be used as an indicator of lexical change (Schøsler, 2007). Indeed, the data seem to show that there is variation in preposition usage, including the choice to not include a preposition, namely to use transitive *andar*. Table 4 shows the cases of intransitive *andar* that appeared with a preposition in each country along with the number of cases of transitive *andar* (i.e. no preposition). Here, there is a high degree of variation among the inner four Central American countries, as seen in the competition between the lack of a preposition and the use of *con* ‘with’. In the countries where transitive *andar* is not widely attested, there is still variation, but variation does not always indicate language change the way language change requires variation. Here, there is notable overlap between the use of *con* and that of other prepositions, but little to no competition occurring with the transitive use, suggesting a lack of lexical shift in these particular subsets of the data. In Costa Rica, El Salvador, Honduras, and Nicaragua, where *andar* has experienced increased grammaticalization (Rodríguez, 2019), it seems that preposition variation has also been restricted, meaning that if a speaker were to use transitive *andar* with a preposition, it would almost always be *con*. In other words, the preposition *con* is not required to produce a transitive meaning.

Country	No preposition (transitive)	con	Other (de, en)	Total
Argentina	0 (0%)	31 (89%)	4 (11 %)	35
Chile	3 (4%)	68 (93%)	2 (3%)	73
Costa Rica	29 (45%)	22 (36%)	10 (19%)	61
El Salvador	25 (63%)	13 (33%)	2 (4%)	40
Spain	2 (6%)	19 (55%)	12 (39%)	33
Guatemala	4 (21%)	14 (74%)	1 (5%)	19
Honduras	26 (53%)	15 (31%)	8 (16%)	49
Mexico	6 (19%)	31 (54%)	17 (28%)	54
Nicaragua	24 (56%)	15 (35%)	4 (9%)	43
Panama	0 (0%)	7 (70%)	3 (30%)	10
United States	2 (5%)	29 (76%)	7 (19%)	38

Table 4: Preposition variation by country

4.3 Competing verbs

Lastly, we took a subset of the data described above (52 tokens), namely the tokens of *andar + dinero/arma(s)* in the four inner Central American countries, as they have the highest frequency in our dataset. We then extracted similar tokens (211) from the Web/Dialects sub-corpus of the CdE (Davies 2016-) with the verb *llevar* to understand whether *andar* overlaps

with other verbs used to indicate alienable possession. This data set was comprised of 263 tokens.

4.4 Results of competing verbs

The results of this comparison can be seen in Figure 3 where the light gray reflects the frequency of *llevar*, which is high, and the black reflects the frequency of *andar* in each country, which is low. In Figure 1, the high p -values indicated a small difference in the distribution of transitive *andar* and non-transitive *andar*, suggesting competition between these two variants. However, here in Figure 3, transitive *andar* does not seem to contrast with *llevar*, at least insofar as overall frequency is concerned. We maintain that this is due largely to the fact that *llevar* is simply more frequent. Chi-squared tests revealed that the differences in frequencies of the two verbs were statistically significant, meaning that the frequencies are so statistically different that it is unlikely that they are competing for the same meanings. Further analysis could reveal that this is due to the choice in nouns, being that *arma(s)* may be more frequent with a verb like *cargar* and *dinero* with a verb like *traer*, both of which are semantically primed for nouns that are compatible with a meaning of alienable possession (e.g., ‘gun’ and ‘money’). However, what is relevant here is that there is a type of layering of meaning occurring in these dialects, which is characteristic of forms undergoing lexical change.

Figure 3: Frequency of Transitive *andar* and *llevar*

To summarize, we maintain that transitive *andar* competes with non-transitive *andar* in Costa Rica, El Salvador, Honduras, and Nicaragua. Here, Polarity, specifically an affirmative reading, proved to be a significant predictor for the use of transitive *andar*. There is, however, no overt competition in our data between transitive *andar* and *llevar* in these four Central American countries where the transitive use of *andar* is most widely attested.

5. Conclusions and discussion

Based on the results of the quantitative corpus analysis, we conclude that *andar* is indeed used with the transitive sense described by Lipski (1994) and Hernández (2002) particularly in varieties of Central American Spanish, namely Costa Rica, El Salvador, Honduras, and Nicaragua. Moreover, its transitive use has become a common way in some varieties to

communicate the idea of ‘to wear or carry on one’s person’ or even ‘to have’ with [+transient possession] nouns. The results of a GLMM point to Country and Polarity as the best predictors for the transitivity of *andar* with the nouns selected for the corpus search.

In comparing the results of the mixed-effects model with those of the hypothesized competition between *andar* and *llevar*, we further conclude that transitive *andar* is not in direct competition with *llevar* specifically, though a layering of meaning is observed both in our data and diachronically (Hopper and Traugott, 2003). However, it clearly overlaps with the intransitive canonical form followed by the preposition *con*. Speakers and writers from these varieties of Spanish use this structure with a possessive meaning, marking for the most part alienable possession. And while the data are limited at this point, transitive *andar* appears to be available to mark inalienable possession, occurring with nouns like *bebé* ‘baby’ or *novio/a* ‘boyfriend/girlfriend’. The data analysis thus far has not produced tokens of transitive *andar* (i.e. without *con*) with complements that are explicitly intransient—a hallmark of prototypical inalienable possession—or perceived as permanent, such as *esposo/a* ‘spouse’ or *hermano/a* ‘brother/sister’. However, abstract nouns such as *año* as complements occur with transitive *andar*, much in the same way that they do with *llevar* ‘to wear/carry’ or even *tener* ‘to have’. The exact role of these complements in the development of *andar* is left for future study.

These results are provided in the context of a proposed case of lexical shift that involves both the gradual diversification of the collocation profile of *andar* and the subsequent entrenchment and expansion of the meaning that arises with the structure *andar con* (Rodríguez, 2019). Here, there is a loss of compositionality and analyzability leading to lexical change wherein *andar con* no longer signifies “to walk with” but “to carry on one’s person.” It is also in this stage where *andar con* appears to take on the properties of a transitive phrasal verb. An

additional step involves the gradual, though not necessarily complete, loss of the preposition *con* with expanded use of *andar* as a type of alienable possession similar to *tener* ‘to have’. Beyond this, we have observed initial evidence that *andar* in Central American Spanish can be used with animate complements to produce a meaning akin to inalienable possession. The extent to which transitive *andar* is employed with a broader range of possessive uses is one of a number of interesting questions left open in this study.

The thread running throughout this analysis is the use of large-scale corpora to uncover patterns in low-resource dialects. This approach relies heavily on access to reliable data that can be understood to reflect, at least in general terms, the language used in particular varieties of the language. Use of the *Corpus del Español* (Davies, 2016-) and other large corpora, such as esTenTen18 (Kilgarriff *et al.* 2014), for tracking lexical change allows for reproducible results that reflect broad patterns of linguistic change, provided (a) that the research questions need not be tied specifically to the demographics of the speakers/users and (b) that the scale of the data samples and subsamples be understood as a countermeasure to issues of representation. Indeed, the claims presented in this analysis would most certainly be even more robust if an exhaustive extraction of *andar* were conducted across each of the large corpora mentioned here (i.e. the *Corpus del Español*, CORPES, esTenTen18), though we will leave this task for future research. Finally, it perhaps goes without saying that a more detailed understanding of lexical change in varieties of Central American Spanish requires an understanding of the socio-historical contexts in which users produce and shape language, factors that are difficult to operationalize in a corpus of 2.1 billion words. But the Big Data approach does help to introduce questions that, in the absence of suitable resources for under-represented varieties, frequently go unanswered or completely unexamined.

Notes

1. The data from Twitter has been revised for ease of understanding. Also, the text has been replicated in the example along with the user handle. All tweets cited in this work were extracted using San Salvador as a target for geo-tagged tokens.
2. SUBES is a type of credit card. We also thank one of the reviewers for a number of insightful suggestions regarding the translations of the examples.
3. We are not discounting the possibility of a null preposition in examples like (16b) and (17b).
4. The notion of inalienable possession as it applies to examples like (2), (14), and (14), while not crucial for this analysis, nonetheless deviates from that described by Zeshan and Perniss. Unlike the relationship indicated in a prototypical case of inalienable possession such as *la mano de Jules* ‘Jules’ hand’, which is permanent, the relationship indicated by *novio* ‘boyfriend’ in example (2) is certainly not permanent in the same sense.
5. The compatability of *andar* with impersonal and passive structures in Spanish is somewhat unclear. A full explanation of the results with *se* passives (e.g., *se anda una camisa roja* ‘one wears a red shirt’) and *ser* passives (e.g., *?la camisa roja fue andada por Dani* ‘The red shirt was worn by Dani’) is required.
6. We follow Rodríguez-Puente who describes phrasal verbs as “lexicalized combinations of a verb and a particle which function semantically and syntactically as a single unit” (2012:71).

References

- Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bybee, J.L. (2008). Mechanisms of Change in Grammaticalization: The Role of Frequency. In B.D. Joseph & R.D. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 602–623). Blackwell Publishing Ltd.
- Bybee, J.L. (2011). Usage-Based Theory and Grammaticalization. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Grammaticalization*, (pp. 69–78). Oxford University Press.
- Company Company, C. (2008). Gramaticalización, debilitamiento semántico y reanálisis. El posesivo como artículo en la evolución sintáctica del español. In J. Kabatek (Ed.), *Sintaxis histórica del español y cambio lingüístico: nuevas perspectivas desde las Tradiciones Discursivas.*, vol. 31 (pp. 17–51). Iberoamericana.
- Davies, Mark. (2002-) *Corpus del Español: 100 million words, 1200s-1900s*. Available online at <http://www.corpusdelespanol.org/hist-gen/>. (Historical / Genres)
- Davies, Mark. (2016-) *Corpus del Español: Two billion words, 21 countries*. Available online at <http://www.corpusdelespanol.org/web-dial/>. (Web / Dialects)
- Evans, N. & Wilkins, D.. (1998). *The knowing ear: An Australian test of universal claims about the semantic structure of sensory verbs and their extension into the domain of cognition*. (Arbeitspapier 32, NF.) Cologne: Institut für Sprachwissenschaft.
- Heine, B. (1997). Grammaticalization theory and its relevance to African linguistics. In R. Herbert (Ed.), *African linguistics at the crossroads: Papers from Kwaluseni* (pp. 1-15). Köppe.

- Heine, B. (2002). On the role of context in grammaticalization. In I. Wischer & G. Diewald (Eds.), *New Reflections on Grammatilization* (pp. 81-101). John Benjamins.
- Heine, B. & Narrog, H. (2009). Grammaticalization and Linguistic Analysis. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 401–423). Oxford University Press.
- Hernández, J.E. (2002). Accommodation in a dialect contact situation. *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 28(2). 93-110.
- Hopper, P. (1991). On some Principles of Grammaticalization. In E.C. Traugott & B. Heine (Eds.), *Approaches to Grammaticalization* (Typological Studies in Language), vol. 19. (pp. 17-35). John Benjamins.
- Hopper, P. & Traugott, E.C. (2003). *Grammaticalization*, 2nd Ed. Cambridge University Press.
- Howe, C. (2018). Grammaticalization. In K. Geeslin (Ed.), *The Cambridge Handbook of Spanish Linguistics* (pp. 604-624). Cambridge University Press.
- Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý y V. Suchomel. (2014). The Sketch Engine: Ten Years on. *Lexicography* 1, 7-36.
- Lehmann, C. (1998) *Possession in Yucatec Maya: Structures – Functions – Typology*. Munich: Lincom Europa.
- Levin, B. & Rappaport-Hovav, M. (1995). *Unaccusativity. At the Syntax-Lexical Semantics Interface*. MIT Press.
- Lipski, J.M. (1994). *Latin American Spanish*. Longman.
- Lipski, J.M. (2008). *Varieties of Spanish in the United States*. Georgetown University Press.
- Mahoney, K.D. (2002). ambulare. *Latdict*. <http://www.latin-dictionary.net/search/latin/ambulare> (21 August, 2017).

- Moreno Fernández, F. (1996). Metodología para el “Metodología del ‘Proyecto para el estudio sociolingüístico del Español de España y de América’ (PRESEEA), *Lingüística* 8, 257-287.
- Moreno Fernández, F. (2005). Corpus para el estudio del español en su variación geográfica y social. El corpus PRESEEA. *Oralia* 8, 123-139.
- Moreno Fernández, F. y Cestero, A.M. (2020). El Corpus PRESEEA. Desarrollos analíticos. *Verba. Anuario Galego de Filoloxía*, 80, 119-138.
- Perlmutter, D. (1978). Impersonal passives and the unaccusative hypothesis. In J.J. Jaeger, A.C. Woodbury, F. Ackerman, C. Chiarello, O.D. Gensler, J. Kingston, E.E. Sweetser, H. Thompson, and K.W. Whitler (Eds.), *Proceedings of the 4th annual meeting of the Berkeley Linguistics Society* (pp. 157-189). Berkeley: Berkeley Linguistics Society
- Perseus Digital Library*. Ed. Gregory R. Crane. Tufts University. <http://www.perseus.tufts.edu> (accessed August 21, 2017).
- RAE. (2017). *andar*. *Diccionario de la lengua española*. <http://dle.rae.es/?id=2ZGgAPa|2ZIziL7> (21 August, 2017).
- Rivas, P.G. (1987). *La Lengua Salvadoreña*. Second Edition. Ministerio De Cultura y Comunicaciones.
- Roberts, I. & A. Roussou. (1999). A formal approach to “grammaticalization.” *Linguistics*, 37(6). 1011–1041.
- Rodríguez, Shannon. (2019). Transitive *andar*: an emerging verb of possession. University of Georgia Manuscript.
- Rodríguez-Puente, P. (2012). The Development of Non-Compositional Meanings in Phrasal Verbs: A Corpus-Based Study. *English Studies: A Journal of English Language and Literature*, 93(1), 71–90.

- Schøsler, L. (2007). The status of valency patterns. In T. Herbst & K. Götz-Votteler (Eds.), *Valency: Theoretical, Descriptive and Cognitive Issues* (pp. 51–66). Walter de Gruyter.
- Sorace, A. (2000). Gradients in Auxiliary Selection with Intransitive Verbs. *Language*, 76(4), 859-890.
- Torres Cacoullos, R. (2011). Variation and Grammaticalization. In M. Díaz-Campos (Ed.), *The Handbook of Hispanic Sociolinguistics* (pp. 148–167). Wiley-Blackwell.
- Torres Cacoullos, R. (2001). From lexical to grammatical to social meaning. *Language in Society*, 30, 443-478.
- van Gelderen, E. (2011). Valency changes in the history of English. *Journal of Historical Linguistics*, 1(1), 106-143.
- Viberg, Å. (2010). Basic verbs of possession: A contrastive and typological study. *Cognitextes*, 4. <https://doi.org/10.4000/cognitextes.308>
- Zeshan, U. & Perniss, P. (2013). Possessive and existential constructions: Introduction and overview. In A.Y. Aikhenvald & R.M.W. Dixon (Eds.). *Possession and Ownership: A Cross-linguistic Typology* (pp. 1-31). Oxford University Press.