On the development of the interpersonal epistemic stance construction in Spanish: the case of *sabes 'you know'* and constructional variants

Abstract

This study contributes to a better understanding of how constructional models can be applied to discourse-level phenomena, and constitute a valuable complementation to previous grammaticalization accounts of pragmatic markers. The case study that is presented concerns the recent development of the interpersonal epistemic stance construction in Spanish. The central argument is that the expanding use of *sabes* as a pragmatic marker can best be fully understood by taking into account the composite network of related expressions which Spanish speakers have at their disposal when performing a particular speech act. The diachronic analysis is documented with spoken corpus examples collected in recent decades, and is mainly informed by frequency data measuring the productivity, as well as formal properties of the construction and its instances.

Keywords

Interpersonal epistemic stance construction, Spanish, recent language change, constructional change, productivity, structural reduction.

1. Introduction

This article investigates the recent development of the so-called "interpersonal epistemic stance construction" (hereafter IESCx) in Spanish. This construction houses a series of routinized expressions by which the speaker (a) directly addresses the interlocutor using a second person verbal form (hence "interpersonal"- see House 2009), (b) asserts the probability of the described state of affairs being true (thus "epistemic" - see Nuyts 2001), and (c) negotiates his own point of view towards the content of a proposition (hence "stance" - see Kärkkäinen 2003 and Traugott 2014).[1] By virtue of its inherent semantics, the lexical class of cognition verbs constitutes a privileged source with which to fill in this construction (e.g. Brinton 1996, Erman 2001 and House 2009 on *you know*; Dostie and de Sève 1999 on French *tu sais*; Lindström and Wide 2005 on related discourse particles in Swedish).[2] In Spanish, expressions such as *sabes* 'you know', *ya sabes* 'you

[1] Although the various expressions covered by this construction convey a wide spectrum of meanings, ranging from the truly epistemic to the metalinguistic, and often crystallize into conventionalized discourse related chunks, in this article, the term IESCx is used as a kind of umbrella term referring to the semantics of the construction as a whole.

[2] Note that 1st person stance constructions (such as *I think, I mean, I suppose* etc.), being more frequent and prototypical comment clauses, have been the focus of an even greater number of studies (Thompson and Mulac 1991; Brinton 1996; Aijmer 1997; Schneider 2007; Van Bogaert 2009; Kaltenböck 2013, among others). However, 2nd person stance constructions

already know', *como sabes* 'as you know', *sabes lo que pasa* 'you know what happens' and *sabes que* 'you know that' possibly followed by a subordinate clause are particularly frequent constructs. These items have received different terminologies and classifications in the literature: comment clauses (Brinton 2008), Reduced parenthetical clauses (Schneider 2007) or, more commonly, discourse markers (Dehé and Wichman 2010) or pragmatic markers (Kaltenböck 2013). They have been the object of two kinds of studies: those which aim at disentangling their intricate web of meanings and functions, and those which target their status as the outcome of a long process of linguistic change.

Across the bounds of language-specific properties, the development of pragmatic markers has been accounted for by different theories and frameworks such as pragmaticalization, lexicalization, subjectification, and without doubt the most widespread, grammaticalization.[3] Indeed, items such as *you know* or *you see* are said to fulfil the diagnostics of grammaticalization: they are the result of a process of reanalysis, being formally fixed and deprived of their syntactic capacities, "decategorized" and "desemanticized". As a consequence they gain syntactic (and prosodic) independence and expand their scope (for a detailed overview of different criteria, approaches and authors, see Brinton 2008). Remarkably, this historical account has led to different conclusions in the literature on the Spanish cognate *sabes*. Whereas, based on its functional properties, most authors (e.g. Vázquez Rozas 2006; López Serena/ Borreguero Zuloaga 2010; Cuenca 2013) do not hesitate to include *sabes* in the category of grammaticalized discourse markers, Martín Zorraquino and Portolés (1999) are more reluctant, mainly because 2nd person markers are said not to fulfil Hopper's (1991) "freezing condition". The marker *sabes* combines with different argument (*tú sabes, lo sabes*) and adverbial complement types (*ya sabes*), can be negated (*no sabes*), and is used with an assertive or interrogative modality (*sabes/¿sabes?*).

One of the main objectives of this article is to show that the attempt to answer the question whether or not a particular linguistic item, to wit *sabes*, is a grammaticalized discourse marker must take into account important novel insights into the processes of language change. A recent study (AUTHOR) has shown that the functional and formal nature of *sabes* has substantially changed in the last decade of the 20th century and at the beginning of the 21st century, so that it can rightly be defined as a pragmatic marker. However, at

deserve at least as much attention given that, on the basis of their morphology alone, they foreground an extra dimension, namely the aspect of intersubjectivity, and thus the relationship between speaker and hearer.

[3] The notion of "marker" has also been subject of a crossfire of diverging theories and lacks an agreed terminology, definition and classification. In this paper, the notions "pragmatic marker" and "discourse marker" are used interchangeably to refer to highly multifunctional linguistic elements which have a "procedural core meaning" and whose specific interpretation is "'negotiated' by the context" (Fraser 1999: 950). For a thorough overview of this topic, see, among many others, Fraser (1999) and Traugott (2007).

the same time, it appears that in present-day Spanish, the use of *sabes* as a pragmatic marker still alternates with structurally complex expressions. In this paper it is shown that this apparent empirical conflict can best be accounted for by complementing the grammaticalization approach with a constructional network approach.

As a starting point, I posit that the development of *sabes* as pragmatic marker cannot be fully understood without taking into account the composite network of epistemic expressions which Spanish speakers have at their disposal when performing particular discourse related functions. Accordingly, this article pursues a triple goal:

1. at the theoretical level, it argues further in favor of a constructional approach to discourse related phenomena, albeit pointing towards some methodological problems that arise when transposing the construction grammar framework to this context;
2. with regard to the specific case study, it aims to provide insight into the taxonomy of the IESCx in general, and the mechanisms of change which have led towards the entrenchment of *sabes* as a pragmatic marker in present-day Spanish;
3. at a methodological level, it illustrates how processes of recent language change can be studied in a corpus of spoken Spanish.

The remainder of this paper is organized as follows. Section 2 reviews the Construction Grammar framework and its suitability for studies of discourse-related phenomena and language change, and sets out the parameters that must be accounted for by the diachronic analysis of the IESCx. Section 3 argues in favor of studying the diachrony of the 20th-21th centuries in order to get an insight into recent processes of change, and defines the shared properties of Spanish interpersonal epistemic stance expressions. Sections 4 and 5 present the results of the corpus analysis and focus on the development of the general schema in terms of its absolute frequency (4.1) and productivity (4.2), and, at a lower level, on the diachronic (re)configuration of the network of related expressions (5.1) and the mechanisms of change (5.2).

2. Towards a constructional approach to the expression of interpersonal epistemic stance

2.1 Why a constructional approach?

Before proceeding to the data analysis, the added value of adopting a constructional model for the diachronic study of IES expressions needs to be addressed.[4] First and foremost, the constructional approach has been put

[4] The construction grammar framework is built upon the idea that language is made up of form-meaning pairings, or constructions, organized in a network (for a thorough introduction to the field, see Hoffmann and Trousdale 2013). The taxonomic network is bonded by inheritance relations between constructions, defined at different levels of schematicity. Ranging from the highest degree of abstraction to the highest degree of concreteness, a distinction is made between schemas (e.g. the caused-motion schema), subschemas (e.g. the

forward as a highly adequate model to account for language change. However, it is important to bear in mind that a constructional approach is not intended as a replacement of more traditional accounts of pragmatic markers in terms of grammaticalization. Indeed it has been argued that the two approaches, grammaticalization and "constructional change" (or "constructionalization"), are not mutually exclusive but in fact complementary models, which, besides a huge overlap in their theoretical and methodological base, partially foreground different questions (Hilpert 2013; Traugott and Trousdale 2013; Trousdale 2014).[5] Whereas grammaticalization has focused mainly on the development of grammatical categories, also often taking into account contextual factors (e.g. Himmelmann 2004) and the regularity of changes at different levels of language (e.g. Company 2006), the constructional approach focuses on how the internal and external configuration of sets of expressions changes. With regard to the two alternative traditions within which grammaticalization is defined – either as a process of reduction and increased dependency or as a phenomenon of contextual expansion – the constructional approach argues in favor of a combination of both views: a constructional change can involve expansion in the range of use of a construction which, at the same time, can become formally more fixed and reduced and appear as a chunk (Traugott and Trousdale 2013: 147).

In addition, a particularly appealing implication of the constructional view, which is not absent but less prominent in grammaticalization accounts of pragmatic markers, is that the perception of relationships between patterns within a paradigm is part of a competent speaker's knowledge. As a consequence, the usage of one particular item cannot be seen as independent of the functioning of others related to it. This entails that linguistic change does not apply to an individual linguistic item alone, but always involves its network "family members". A construction can thus be assumed to have its own history which is, however, often constrained by, or has an impact on, the history of related patterns or schemas in the network. This interconnectedness can take very different forms and has consequences for both the internal (a)

causative subschema), micro-constructions (e.g. different constructions to express causation) and constructs (the empirically attested tokens). The vertical inheritance perspective accounts for generalizations across constructions, but leaves room for the exceptional behavior of (sets of) members (Goldberg 1995).

[5] In this article, the term "constructional change" is used instead of "constructionalization". Traugott and Trousdale (2013: 22) define constructionalization as the creation of a $form_{new}$/$meaning_{new}$ pairing as a new node in the network. But this study is not involved with the creation of the IESCx as a new construction, a phase which is to be situated in earlier times but is in fact hard to trace. I believe that what we are observing is in fact a series of "post-constructionalization constructional changes" (in terms of Traugott and Trousdale 2013: 27), abbreviated here as "constructional changes". As defined by Hilpert (2013: 16) this can refer to any change with regard to the function, form, frequency, or distribution of the construction.

and external (b) configuration of a network (Hilpert 2013; Traugott and Trousdale 2013):
(a) Within a particular network, new patterns can arise, and older ones can disappear; a pattern can become more frequent and productive and can invade the domain of another, or, conversely, it can lose its vitality, become obsolescent and leave a functional gap.
(b) As a consequence, the configuration of the network itself can remain stable, can come to be reorganized, grow or shrink.

The view taken in this study squares with the recent and growing interest in applying constructional models and insights to the study of discourse-level phenomena. Until the onset of this century, the focus of Construction Grammar (CxG) was mainly on argument structure constructions, but more recently, lower levels (i.e. morphology and phonology) as well as higher levels have been more explicitly integrated into the model. This development is completely in line with the *all-inclusive* principle of CxG, which declines to distinguish between core, peripheral or even ungrammatical patterns (Brinton 2008).

Also non-propositional material typically found in spoken discourse, such as particles (Fried and Östman 2005), pragmatic markers (Traugott in press), parentheticals (Hilpert 2013) and comment clauses (Kaltenböck 2013; Van Bogaert 2009, 2010), has been argued to be part of grammar and of a speaker's constructional knowledge. The central idea is that different expressions with analogous and overlapping functions and formal characteristics are instances of an overarching construction. Thus one of the main advantages of the CxG approach is that it reveals generalizations across particular items. Fischer (2015) shows, for instance, that the formal behavior of the English particle *oh* (defined by the structural contexts in which it occurs) interacts with its meaning (composed of a series of different communicative tasks), and argues that these form-meaning pairings count as schematic structures and can thus apply to other particles as well. In a recent paper, Traugott (in press: 9-10) argues in favor of a general pragmatic-marker macro-schema, defined by a very general form ("position typically precedes or follows the clausal construction it accompanies, or the finite verb of a clausal construction") and meaning ("pragmatic, non-contentful meaning, and scope over the clausal construction it accompanies") components. At the schema level, metatextual, epistemic and interpersonal pragmatic marker Cxs are distinguished, which are then further concretized in different discourse marker subschemas.

Whereas the idea of a constructional approach to pragmatic markers and associates is in itself perhaps not so novel, to date, not much work has been done on their history within the constructional change framework. This, of course, leads us directly to the main research question of this paper and the aim of accounting for the apparently contradictory tendencies observed in the literature (see Section 1), namely that: (a) *sabes* increasingly acts as a

(grammaticalized) pragmatic marker but (b) at the same time still alternates with a series of complex alternative expressions.

2.2 Parameters of analysis

It has been observed that constructional changes involve different phenomena. Traugott and Trousdale (2013) emphasize the expansion of the productivity and schematicity of constructions and the reduction of their compositionality. Expansion of productivity mainly refers to a grammaticalizing construction's increase in the range of collocates that co-occur with it. This collocational expansion is also known as "host-class expansion" (Himmelmann 2004), or the increase of a construction's type frequency. Indeed, a higher degree of constructional vitality coincides with a higher probability of the constructional schema generating new instances. This can be seen as a correlate of the increasing degree of schematicity of a construction. When constructions are used more frequently, speakers generalize over the individual types and develop more abstract schemas. In their turn, this abstract schema attracts new types, which again contributes to the productivity at the schema level (Barðdal 2008). In addition, a grammaticalizing construction suffers a gradual decrease in compositionality; that is, the meaning of the individual parts no longer coincides with that of the construction as a whole. Additionally this can lead to formal (morphological and phonological) reduction.

Hilpert (2013) adds that, in order to understand how constructions have merged and developed, it is crucial to observe possible changes in the distribution of different structural variables within a more general schema. These changes can point towards changes in the configuration of a network (see Section 2.1 above), such as the creation or decline of particular structures, or the internal reorganization of a network.

These diagnostics or correlates of constructional change can guide us through the diachronic analysis of the IESCx. However, to my knowledge, no previous study has attempted to empirically operationalize these principles for the analysis of pragmatic markers and related constructions. As a result, a number of methodological (and terminological) questions are raised during the analysis. Among these are: how can we account for the productivity of the IESCx? What is its type frequency? How can its degree of compositionality and formal reduction be accounted for? Indeed, although an increasing number of scholars have argued in favor of a constructional approach to discourse-related phenomena, its implementation still needs to be worked out in more detail. This study can be considered as a first step in that direction.

3. Defining the interpersonal epistemic stance construction

3.1 Data collection

This study starts from the assumption that language change is not restricted to drastic long-term shifts affecting the language system as a whole; it can

also consist of a more subtle variation in language use that comes about in the course of just a few decades. The study of these "current changes" defined as "changes in the language that have taken place over relatively short spans of time, over decades rather than centuries" focuses on changes in the recent past, and often also in contemporary language (Aarts et al. 2013:1). As is the case for any data-driven investigation, good sampling of data is crucial - and, alas, not a walk in the park.

First, the monitoring of the use and development of the IESCx can only be achieved on the basis of a spoken data corpus, not only because spoken language is known to be "at the forefront of linguistic innovation" (Kaltenböck 2013: 287), but mainly because the construction under scrutiny tends to be exclusively used in informal conversational language. But, for obvious reasons, it is difficult to have access to a non-contemporary native speaker's linguistic intuition (Jucker 1995). Second, short-term patterns of change can only been detected when differences in the corpus cannot be attributed to factors of influence other than the time variable.

Both restrictions constitute a methodological challenge for the present purpose, given that spoken corpora for Spanish are not as widespread as for other languages such as English.[6] However, I have maximally striven for a large corpus including reliably transcribed texts, and, most importantly, with data sampled from different decades. I have resorted to a "micro-diachronic" analysis of the development of the IESCx over four decades, starting with the period for which the first spoken corpus for peninsular Spanish is available, namely the 1970s.

The database was compiled from eight existing spoken language corpora. It includes interpersonal epistemic stance (IES) expressions which all appear in a dialogic context of spontaneous conversations. These are face-to-face conversations (e.g. COLAm, CORLEC, C-Oral-Rom), telephone conversations (e.g. CREA oral), interviews in television and radio shows (CREA oral), and semi-directed interviews (e.g. Habla culta de Madrid, COSER). The tokens were selected through a lexical query for the string *sabe(s)*, and were then manually sorted in order to discard full lexical uses of the verb. These include all instances of *sabes* followed by a Direct Object with full lexical meaning (1a-1b), in which the verb directly refers to the knowledge of the interlocutor, without negotiating any point of view towards the contents of the message. The DO can then be an NP (as in 1a) or a subordinate clause (as in 1b).

(1) a. Suponemos que te *sabes* la pregunta que hemos hecho para concursar. ('We suppose *you know* the question that we have asked in order to be able to participate') (CREA Oral, 1991)

b. ¿Qué cara pones tú cuando *sabes* que tu mujer le gusta alguien, o que tiene un amante […]? ('How would you react when you got

[6] It should however been specified that compared to other Romance (e.g. French or Italian) and Germanic languages (e.g. Dutch or Swedish), spoken Spanish is better represented.

to *know* that your wife likes someone, or that she has a lover.') (CREA oral, 1996)

These uses contrast sharply with the intersubjective stance meaning of the IES expressions included in the database. In (2a), *sabes* is followed by the semantically vague NP *una cosa*, and the expression as a whole has the main function of calling upon the interlocutor's intention, just like the pragmatic marker *sabes* further in the context does. The same applies to (2b) in which *sabes* followed by the subordinate DO *que pasa* fulfils a mainly phatic function (see also Sections 1 and 3.2 for more a more detailed definition of the IESCx).

(2) a. […] ¡ay! dejarla no la obliguéis pobrecita qué se va a beber eso *sabes una cosa* tía es muy pegajoso *sabes* está todo el rato. ('Ay, leave her, do not force her, poor thing, she will drink it. *You know something* girl, it is very tempting, you know, she is here all the time.') (C-Oral-Rom, 2001)

b. MAORE2J01: pero con quien os íbais ('but with whom were you going'?)
MAORE2J02: *sabes que pasa* que he estado todo el di toda la mañana grabando y no ha grabado nada porque está roto tía ('you know what happens that I have been recording the whole morning and I haven't recorded anything because it is broken, girl.')

The result is a large database of 3279 tokens, representing speech of both male and female speakers, of mostly intermediate sociocultural level (although higher and lower levels are also included), and of different generations (e.g. adolescents in COLAm, elderly in COSER) living in Spain.[7]

Table 1 gives an overview of the different corpora that have been consulted for each decade and the number of IESCxs attested in each.[8] In a last phase,

[7] With regard to the data collection, three caveats should be mentioned. First, note that the different sociolinguistic attributes can also have important explanatory power, especially age-related classification. Teenage talk (COLAm) is indeed said to play an important role in linguistic change (Zimmerman 2002). Besides, teenage talk data can be very good indicators of how the use of a linguistic phenomenon has gradually spread over time. An alternative methodology would be to engage in an "apparent time analysis" (the synchronic study of age differences) with respect to some linguistic phenomenon (Cameron 2011). In future research, the outcomes from this diachronic study will be compared to such a study. Secondly, a conscious decision was made to restrict the data to peninsular Spanish. It is highly probable that other dialects of Spanish resort to other lexical elements to express similar functions, such as e.g. *¿cachái?* in Chilean (Urzúa-Carmona 2006). For practical reasons, however, this diatopic factor does not serve as a quantitative research parameter in this article; the sociolinguistic analysis will constitute the topic of a separate study. Third and unfortunately, for the 1970s and 1980s I only have access to a restricted set of spoken language data. However, along with Nevalainen and Raumolin-Brunberg (2003: 26), I believe that we must "make the best use of the data available" bearing in mind that the description of the development of the IESCx can be limited by our data.

[8] See Enghels, Vanderschueren and Bouzouita (2015) for a more detailed description of the features of each individual corpus. The corpora were consulted in the summer of 2015.

these hits have been manually annotated for a series of functional and formal properties which are detailed and explained throughout this article.

Table 1. Overview of corpora and data extraction

decade	corpus	corpus size	# IESCx
1970s	Habla culta de Madrid	ca. 150,000	72
1980s	CREA oral	not defined	61
	Val.Es.Co	ca. 15,000	19
1990s	CREA oral	not defined	359
	CORLEC	ca. 269,500	83
	Val.Es.Co	ca. 251,900	63
	Coser	ca. 950,000	516
2000s	Val.Es.Co	ca. 2000	3
	Coser	ca. 730,000	503
	C-Oral-Rom	ca. 300,000	215
	PRESEEA	ca. 216,046	118
	COLAm	ca. 456,000	1266
Total			3279

One of the main characteristics of a construction is its non-compositional nature (Traugott and Trousdale 2013). A construction appears as a chunk, following a mismatch between the meaning of the individual elements that make up the construction and the meaning of the construction as a whole. This characterization applies in varying degrees to the following expressions found in the database, ordered from most to least frequent: *sabes* (n=2385)[9], *sabes lo que V* (251), *ya sabes* (132) / *sabes ya* (1), *sabes que + clause* (69), *tú sabes lo que V* (55) / *sabes tú lo que V* (16), *tú sabes + cl* (52) / *sabes tú + cl* (3), *tú sabes* (46) / *sabes tú* (5), *ya sabes + cl* (42), *no sabes* (39), *como tú sabes* (18) / *como sabes tú* (2), *no sabes lo que V* (19), *sabes qué* (18), *ya sabes lo que V* (15), *como sabes* (13), *no lo sabes* (11), *ya lo sabes* (10), *lo sabes* (10), *tú no sabes* (10), *tú no sabes lo que V* (6), *tú lo sabes* (6) / *lo sabes tú* (2), *ya sabes tú* (4) / *tú ya sabe*s (2) / *ya tú sabes* (1), *tú ya sabes + cl* (4) / *ya sabes tú + cl* (4), *tú no sabes + cl* (4), *como sabes + cl* (3), *sabes (una) cosa* (3), *tú sabes + SN* (2), *ya sabes + cl* (2), *como sabes lo que V* (1), *te lo*

[9] Note that in this article *sabes* is consistently referred to without question marks representing the interrogative modality. An attentive reviewer found it advisable to better distinguish between an interrogative *sabes* and an assertive one. Despite of the validity of this observation, the issue is not easy to deal with due to different transcription norms of the corpora listed in Section 3.1. Most corpora consistently present *sabes* between question marks (e.g. C-Oral-Rom, CORLEC, Coser,…) independently of its prosody, whereas others never do (e.g. COLAm). Consequently, the punctuation does not seem to offer a correct image of the declarative vs. interrogative modality of the marker. The only solution would then be to proceed to an acoustic analysis of the marker, but this, besides going beyond the limits of this paper, is also hard to realize given that not all recordings have been made publicly available.

sabes (1), *como ya sabes + cl* (1), *tú ya sabes lo que V* (1) / *ya sabes tú lo que V* (1) / *ya tú sabes lo que V* (1), *no lo sabes tú (1)*, *tú sabes lo que V* (1), *no sabes + cl* (1), *no sabes tú* (1), *no sabes tú lo que V* (1), *no sabes tú + cl* (1), *sabes cómo V* (1), *tú ya sabes algo* (1).

These instances can be seen as (part of) the "cloud of exemplars" (Bybee 2010: 14) that represent the IESCx in the Spanish speaker's mind.[10] This cloud, in fact, represents constructions which can be placed on a continuum of schematicity, going from the most abstract macro-construction to its most concrete instantiations or constructs.

3.2 Functional definition of the interpersonal epistemic stance construction

As shown by the above-mentioned list, speakers have at their disposal many different expressions in order to convey personal feelings, value judgements, assessments or attitudes towards things, beings, processes or facts (Biber et al. 1999). With regard to 2nd person expressions with nucleus *sabes*, the complex notion of "stance" can be reduced (following the terminology of Brinton 2008) to three main functions, to wit, an interpersonal, modal and textual one.

In order to avoid an intuition-based classification, it is crucial to come to an operational definition of these pragmatic functions, open to objectification (as has been argued for especially in recent studies on discourse-related phenomena such as Torres Cacoullos 2007, Hernández and Baldazo 2013 or Tanghe 2016a, 2016b, but also in former approaches such as Schwenter 1999). This goes in line with the general tendency in contemporary linguistic research to maximally strive for empirical and quantifiable methods in the research on semantic and pragmatic meaning (e.g. Enghels and Jansegers 2013; Glynn and Fischer 2010). However, I agree with Geeraerts (2010) who states that the best way to proceed is to combine different methodologies (which then engage in the so-called 'empirical cycle'), taking into account that not all aspects of the meaning have a direct reflection on the morphosyntactic or distributional behavior of linguistic items. Moreover, although they frequently go hand in hand, objectifiable criteria not necessarily equal quantifiable criteria, as for instance a particular pragmatic function can be observed through a particular co-occurrence pattern (as is shown below), without it being possible to integrate that pattern as a (quantifiable) variable in a more general model.

In concrete, in order to define the macro functions of the IES expressions, each occurrence has been subjected to a detailed analysis of its contextual behavior. The classification of the functions has been decided on, taking into account three complementary parameters:

[10] Of course, other second person expressions derived from cognition and perception verbs enter the domain of interpersonal epistemic stance. They include *entiendes* ('do you understand'), *(ya/como) verás* ('as you will see'), *fíjate* ('imagine'), as well as pragmatic markers such as *bueno* ('well'). The tracing of the larger scale network goes beyond the limits of this article and is a subject for future research.

1. the pragmatics and meaning of the message, which can be neutral (mainly with the interpersonal function), or can be marked by an explicit positive or negative prosody (mainly with the modal function, as illustrated below). A series of paraphrase tests help to define this semantic prosody in a more precise way;
2. recurrent co-occurrence patterns observed in the data, which range from explicit references that the speaker makes to him or herself (modal function), to particular reactions on behalf of the interlocutor (interpersonal function). While the observation of these patterns certainly provides an independent basis for the functional classification, the patterns themselves can hardly be quantified. That is, the absence of a particular collocational pattern does not automatically preclude a particular functional classification.[11]
3. the position of the stance expression, which can be related to its pragmatic function on a quantifiable basis (following previous accounts of the form-function relationship of pragmatic markers, such as Beeching and Detges 2014; Degand and Fagard 2011; House 2009; Tanghe 2016a, 2016b amongst many others). Given that the numerous studies recently conducted on the relationship between the position of pragmatic markers and their functions have analyzed diverse kinds or markers in various languages, they have sometimes led to different conclusions. However, all sources agree on a form-function correlation, although not on a one-to-one basis. The analysis of the position of the IES expression relies on an economized version of the Val.Es.Co model of discourse segmentation (Briz and Bordería 2010), and distinguishes between the position of the expression at the level of the interventions or turns of the speakers, and the turn-internal level. With regard to both discourse units, a pragmatic expression can be placed at the beginning, in the middle or at the end; or, it can appear in an independent position.

In the dataset, the three macro functions performed by IES expressions are rather evenly distributed: interpersonal function (35,3% or 1157/3279), modal function (38% or 1246/3279), and textual function (26,7% or 876/3279). Let us now turn to the more detailed definition of these three domains.

To begin with, an IES expression can convey an interpersonal function, mainly centered on the interaction the speaker wants to establish or maintain with the interlocutor. This function mainly coincides with the expression being placed at the extreme left (3b, 3c) or right (3d) periphery of a speaker's turn (in 47.5% of all cases, 550/1157), or at least at the left (4a-b) or right (4a) periphery of a particular utterance within a turn (another 42.5%, 492/1157). This macro function encloses two subtypes, namely the epistemic and interactional functions. First, as an epistemic marker, the IES expression is an interactional marker mainly directed toward the interlocutor, with whom the speaker wants to share responsibility for the correctness of the information

[11] To compare, the absence of a particular pattern in a corpus cannot automatically be held as an argument against the existence of the pattern in language as a whole.

being communicated. The paraphrases would then be the affirmative 'I know you share this knowledge with me' (3a, 3c, 3d), or interrogative 'Do you share this knowledge with me' (3b). This function coincides with the presence of contextual indices of interaction such as direct references to the addressee (e.g. *¿Te interesa eso?* in 3b), or an affirmative or negative answer on behalf of the addressee (e.g. *Gracias, pero…* in (3a), *ah, sí* in (3b), *sí* in (3d)). In these contexts the construction conveys a primary topical function (3a, 3c), but it can also be used to introduce new information, presented as possibly shared between the interlocutors (e.g. the reference to the price of the object in Portugal in 3b).

(3) a. <H1> No sé, no sé. No sé, no sé. Pero bueno, que... si vienes y si te apetece, pues... ('I don't know, I don't know. I don't know, I don't know. But ok, that… if you come and you like it, then…')
<H3> Qué? ('what?')
<H1> Que si vienes y te apetece, *ya sabes*. Se abre el día 1 de junio. ('If you come and you like it, *you know*. It opens on the first of June.')
<H3> Gracias, pero yo empiezo con los ex menes. ('Thank you, but I start with the university exams')

b. H1> Muy bonito... Que es un violetero; yo no sabía que era un violetero. Pero lo es. ('Very nice… It is a vase with violets; I didn't know it was a vase with violets. But it is')
<H3> ¿Te interesa eso? ('Are you interested?')
<H1> Si es precioso, pap, para adornar! ('yes, it is gorgeous, dad, to decorate')
<H2> ¿*Sabes*? En Portugal valían... los ratoncitos como el vuestro y eso... Valían cada uno cinco mil <simult neo> pelas... ('*You know*? In Portugal they would cost … the little mice like yours and that … they would be five thousand each…')
<H3> ¡Ah, sí! ('Ah, indeed!') (CORLEC 1991)

c. E: *Como ya sabes que* hay algunos pueblos que tienen nombres raros... ('*As you already know,* that there are some villages that have strange names') (Coser, 1999)

d. A: Seguramente para hacerte de oro, te iba a decir. Es que el silencio da mucha personalidad, *ya sabes*. ('Certainly to make you successful, I was going to say. Silence gives one a lot of personality, as you know.)
B: Sí. Oye, pues por eso ('Yes. Listen, well that's the reason.') (CREA Oral, 1988)

Secondly, besides the epistemic meaning, the interpersonal macro function also groups numerous cases in which the IES expression is used to control the interaction between speakers. This function also comes in at least two varieties. When adopting a phatic function, the IES expression verifies whether the interlocutor is handling the flood of information ('are you still

following me?'). This is frequently found in contexts where the speaker sets up a rather large monologue on a particular topic (4a). When adopting a conative function, the IES expression asks for collaboration or an intervention on behalf of the interlocutor (e.g. the addressee has to pass something on to the speaker in 4b). The interpersonal function is supported by the affirmative answer in (4b).

(4) a. Exacto, ya yo les he explicao que últimamente hacía con puerro y cebolla, porque los de Mondragón… *Ya sabes*, mi primo y todos estos, pues estos allí han acostumbrados a la morcilla de puerros y cebolla, y entonces pues la cebolla se refríe con la manteca del cerdo. ('Indeed, I have already explained you the recently I made it (the lard) with leek and onions, because the ones of Madragón... *You know*, my cousin and those people have become used to sausage with leek and onions, and then the onions are cooked with the lard of the pork.) (Coser, 2000)

b. <H1> Eso está bien. Jo, además eso a mí también me interesa. ('That is good. Moreover, I'm also interested')
<H2> Pues sí... es verdad, como te interesa a ti también... Y si pides presupuesto a Inés y te los compra, pues *ya sabes*, me los pasas. ('Okay, it's true, you are also interested… And if you ask Inés an estimate, and she buys them for you, you know, give them over to me.')
<H1> Vale. <risas> ('Okay.' <laughs>) (Corlec, 1991)

Next, IES expressions have a more subjective and modal macro function when they are used by the speaker to reinforce the expression of his/her own emotions or opinions, which the interlocutor can agree with or not. Hence the interpersonal function is not completely absent, but backgrounded. In the corpus this function is predominantly expressed by IES expressions in turn-medial position (70% or 872/1246) (5a-5c). Specifically, they allow the speaker to reveal a personal interpretation of the information by emphasizing some part of an utterance (namely the positive evaluation of the way of life of a third person in (5a), or good way of life of the speaker in (5c)), or by expressing (dis)content (5b). In the corpus, this function coincides with the presence of exclamation marks, recurrent references to the 1st person singular (as in 5b-5c), semantically loaded positive or negative words (as for instance *consuelo de tontos* in 5b), or the use of intensifiers (e.g. *lo bien que* in 5a).

(5) a. Pero además *no sabes tú lo* bien *que* vive, maja, porque le tiene al marido fuera toda la semana ('But what is more *you don't know* how well she lives, girl, because her husband is away all week'). (CREA oral, 1991)

b. I: sí / exacto me dice porque si es un es un trabajo bueno por eso te lo han copiado/ si no fuese bueno no te lo hubiesen copiado y yo sí pero eso es consuelo de tontos *¿sabes*? ('indeed, he tells me it is good work therefore they have copied you, if it weren't good

they wouldn't have copied your work, and I okay but that's cold comfort *you know*') (PRESEEA 2001)

c. I: sí un frío pero bueno yo creo que hay que tener también medidas además que hay que <vacilación/> bueno hay que tener medidas *¿sabes qué pasa?* que yo tomo todos los días naranja por la mañana ('Yes a cold but ok I think one should take some actions and that you <hesitation> well you should take some actions. *Do you know what happens?* That I eat an apple every day in the morning.') (PRESEEA, 2001)

Finally, the IES expressions can also operate at the textual level when they are used as discourse connectors coordinating the individual parts of a conversation in order to maintain coherence. Just like IESs with a predominantly modal function, the expressions adopting the discourse organizing function are mostly found in the corpus in turn-medial position (83,2% or 729/876). But what is different, is that the IESs then typically interrupt (6a) or even distort the syntax of the utterance (in 6b-6c the expressions *sabes* and *como sabes* are inserted right in the middle of an utterance). When adopting this function, they allow the speaker to introduce a new discourse unit that specifies or explains a previous one (specifying the positioning of the dog in 6a), or to reorient aforementioned information (6b). They can also be used as gap-fillers (to be paraphrased by other fillers such as *uhm*) which provide the speaker with some extra time to look for words or to plan the upcoming message while avoiding silence (6c). In the dataset, the textual function is revealed by the following diagnostics: presence of repeated words or parts of the utterance (in 6b *cuando yo cuando yo*), reformulation or specification of parts of an utterance (6a), the use of vague words (y *etcétera, etcétera* in 6c), or repeated pauses (6c).

(6) a. CLE: […] mira le ha llevado Vidal esta mañana por el lago o por ahí y viene ya cansado y se tumba debajo del sofá como siempre *ya sabes* entre los dos sofás […]. ('Look Vidal took him to the lake this morning, and he arrives here very tired, and he lies down under the couch as usual *you know* between the two couches') (C-Oral-Rom, 2001)

b. MAESB2J01: es que joder cuando yo cuando yo me voy de campamento no es que *sabes*/ igual cogemos un autobús que sale a las doce de la noche y llegas allí a las cinco […] ('It is that damn I when I go camping it is not that *you know* we just take a bus that leaves at 12 o'clock at night and you get there at 5') (COLAm)

c. I1: de que ella... le | estamos en la casa con ellos y caseros y le *como sabes* y etcétera, etcétera. ('that she… her | we are in the house with them and the landlords and her *as you know* and etcetera etcetera') (Coser, 2009)

In sum, this overview clearly demonstrates that different stance expressions share functional behavior. This motivates their common classification under

the head of a more general abstract schema or construction, namely the IESCx, whose structural characteristics are further specified in Sections 4 and 5 below.

4. General productivity measures of the IESC schema

That frequency data provide strong evidence for the verification of linguistic hypothesis in general is now widely accepted in the literature (see, for instance, the edited volumes by Gries and Divjak 2012 or Behrens and Pfänder 2016). Their validity for the exploration, description and motivation of constructional changes has also been borne out (e.g. Bybee 2003; Torres Cacoullos and Walker 2011). Hilpert (2013: 207) even argues that, within a usage-based exemplar model of grammatical knowledge, frequency measures give direct access to linguistic knowledge. A frequency change of, and within, a cloud of exemplars can be indicative of processes of language change at a higher level. However, the notion of frequency in itself is multidimensional as it relates to different research questions. Two kinds of measurements are reported below: the evolution of the token frequency (4.1) and type frequency (4.2) of the construction at its most schematic level.

4.1 Token frequency of IESCx

Token frequency has unequivocally been defined as the frequency of occurrence of a linguistic entity (be it a morpheme, a lexeme, a string or a construction) in running text. A high and increasing token frequency of an item would be both "a concomitant of its grammaticalization" (Thompson and Mulac 1991: 319) and a trigger for upcoming formal and functional changes of the item (Bybee 2003: 605).[12] As a first encounter with the data, Table 2 indicates the absolute (n) and normalized (N, tokens per 10.000 words) frequencies (%) of the IESCx per decade. These figures have been calculated by merging the results for the various corpora available per decade (see Table 1 above).[13]

[12] Indeed, grammaticalizing items are said to be necessarily highly frequent in speech, but grammaticalization increases their token frequency even more, as their use expands to more contexts.

[13]For the 70s the only available data come from the *Habla Culta* corpus; for the 80s I only have access to CREA Oral. Given that the CREA corpus does not provide data on the total amount of words for the different subcorpora, the normalized frequencies cannot be calculated (hence the indication [?]). A part of the data coming from CREA oral is also included in the 90s corpus. However, the normalized frequencies are computed taking out this subcorpus, and including only CORLEC, Val.Es.Co and the relevant part of Coser. The other part is included in the >2000 subcorpus, together with the data from C-Oral-Rom, PRESEEA, and COLAm.

Table 2. Token frequency of IESCx

Period	IESCx	
	n	N
70s	72	4.8
80s	80	[?]
90s	1021	4.5
>2000	2105	12.3

It immediately becomes clear that over the last four to five decades the use of the IESCx has increased dramatically: from 4.8 tokens per 10.000 words in the 70s corpus to 12.3 in the >2000-corpus, the most important leap being situated in the new century. The log-likelihood goodness of fit test gives a result that is significant at an error level of $p < 0.001$ ($G2 = 84{,}23$). This suggests that the observed increase in the use of the IESCx between the 1970s and the 21st century is likely to represent a real increase in the population and should not merely be explained by accident.[14]

There may be several causes for this observed increase. A first explanation could be that it is an effect of the nature of the data gathered for each subcorpus, to wit the more formal nature of the 1970s corpus, which includes semi-directed interviews, as opposed to informal spontaneous speech in the 2000s. However, the results from the two corpora providing data for both the 1990s and 2000s, namely Coser and Val.Es.Co (see Table 1), mean that this explanation can be, at best, only a partial one: the Coser values rise from 5.4 to 6.9, and those of Val.Es.Co from 2.5 to 5.6. Another plausible conclusion for the observed increase, then, could be that it is one of the symptoms of the increasing colloquialization of present-day Spanish. Note also that within the 2000s sample, the value for the COLAm corpus (N=27.8) is the highest. This, however, does not need to invalidate the hypothesis that in general the use of the construction has expanded, given that, in an Apparent Time Analysis (cf. footnote 7 above), youth speech represents the most recent language stage.

This drastic numerical increase points towards the conclusion that, with regard to its external dimension, the network has exponentially grown. The IESCx has settled itself more as a schema to which speakers more frequently resort in the flow of discourse, and to which, possibly, more constructs are attracted. This possibility leads us towards a second operationalization of the notion of *productivity*, namely the type frequency of a construction.

4.2 Type frequency of IESCx

The notion of type frequency has received multiple definitions, depending on the linguistic phenomenon being observed. In its broadest definition it can be defined as "the number of items that are available to a particular class of forms" (Hopper and Traugott 2003: 127). In constructional approaches, the

[14] For more background information on this test, see http://ucrel.lancs.ac.uk/llwizard.html (consulted in May 2016).

notion receives particular attention given that it accounts for the expansion or reduction of the linguistic material available for a general schema. The literature itself provides rather vague indications on what type frequency exactly is: Traugott and Trousdale (2013) resort to the notion of "possible collocates" of a construction; according to Bybee (2003: 605) one should count "the different lexical items with which a construction can be used". However, for the study of the IESCx, the measure of type frequency is operationalized following Hilpert's proposal (2013:128), namely by registering the "different instantiations of the same construction". A continuing increase in type frequency over time would then count as evidence of the productivity of a construction (Barðdal 2008; Hilpert 2013; Traugott and Trousdale 2013).

With regard to its formal make-up, the IESCx contains four different slots which can be filled or not. Besides the verbal nucleus *sabes*, there is:

a. a slot making the 2nd person subject *tú* ('you') explicit;
b. two slots for adverbs specifying the mode of knowledge (*como, ya* or *no*) (e.g. *como sabes* 'as you know');
c. a slot for the DO of *sabes* (e.g. as in *lo sabes* 'you know it')

Each time a different item fills in one of these slots, resulting in different lexical combinations, this is counted as a different construction type. Figure 1 details the type frequency of the IESCx per decade.

Figure 1. Type frequency of IESCx per decade

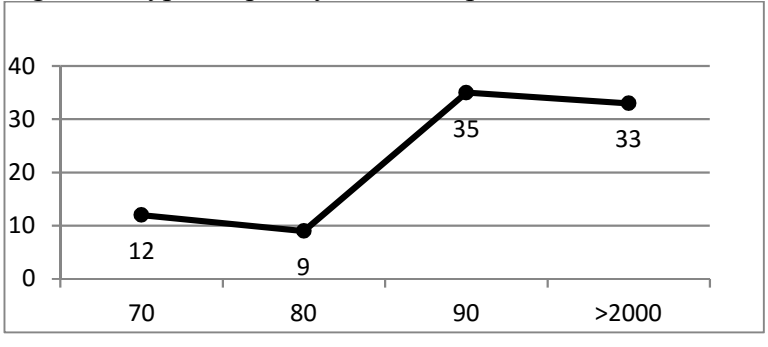


Figure 1 tells us that since the 70s, the type frequency of the epistemic stance construction has evolved following an S-shaped curve with (1) a gradual beginning, with even a little regression from 12 to 9 types in resp. the 70s and 80s corpora, followed by (2) a rapid expansion of the construction to 35 different types in the 90s corpus but (3) a gradual tapering off towards 33 types in the >2000 corpus.

However, given that the type frequency heavily depends on the token frequency of a particular construction in a corpus, the above calculation must be considered a rather naïve account of productivity. More nuanced approaches have been proposed in the literature (see e.g. Zeldes 2012 for an overview) but, unfortunately, they cannot be applied straightforwardly to the Spanish corpora used here, about whose compositional nature hardly any

information is available (e.g. none of the corpora listed in Table 1 is lemmatized or POS-tagged). As an intermediate solution, one can simply calculate the type/token ratio for each decade. The higher the resulting number, the higher the productivity of the construction. This calculation generates a totally different image with regard to the evolution of the construction's productivity:

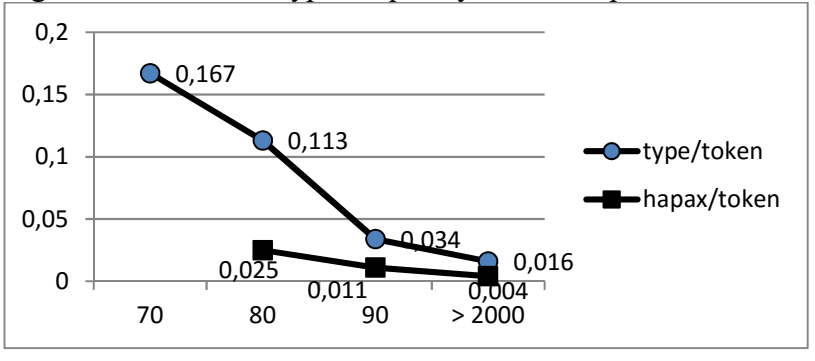
> 1970s: 0.167 (12 types / 72 tokens) > 1980s: 0.113 (9/80) > 1990s: 0.034 (35/1021) > 2000s: 0.016 (33/2105)

Alternatively, as a variant of this measurement, inspired by Hilpert (2013: 128) one can count the *potential productivity* of the construction by dividing the number of hapaxes of the construction (thus, non-previously attested types) by the overall frequency of the IESCx. This gives the following figures per decade starting from the 1980s:

> 1980s: 0.025 (2 hapaxes/80 tokens) > 1990s: 0.011 (11 hapaxes / 1021 tokens) > 2000s: 0.004 (9 hapaxes / 2105 tokens).

Figure 2 plots the two productivity measurements for each decade.

Figure 2. Relativized type frequency of IESCx per decade



As a conclusion, it can be stated that the productivity of the schema itself seems to have dropped considerably. Thus, whereas the construction itself seems to have grown in terms of absolute numbers (see Section 4.1), the figures presented in Figure 2 point towards an internal reorganization within the IESCx network. It is shown in Section 5 that these frequency data indeed reveal that particular subschemas of the construction have marginalized, whereas others have considerably gained in vitality.

5. Configuration of the network: a diachronic survey

As the list above (Section 3.1) illustrates, the IESCx comprises a heterogeneous set of constructional variants (or "allostructions", in Cappelle's 2006 terms) which can be classified according to their degree of structural complexity. On various accounts, it has been argued that the availability of different options for the speaker is an important locus of

language change, and that the changing distribution of constructional variants provides valuable insights into this process (e.g. Cappelle 2006; Hilpert 2013). With regard to the IESCx, the language user's preferences in the choice between variants may not have remained constant over the last 50 years. This section examines the extent to which such a diachronic development of the relative frequencies of structural variants can be observed, and whether some general tendencies can be detected (Section 5.1). It also offers an explanation for the attested changes (Section 5.2).

5.1 Structural reduction within the epistemic stance construction

The IESCx comprises different variants that exhibit a varying degree of finiteness, and that can be ordered on a cline from simple structures (e.g. *sabes*) to fully finite clauses (e.g. *sabes lo que pasa* 'you know what happens'). These structural variants constitute an important variable for the present diachronic empirical research. Abstracting over specific attestations (or constructs), four different patterns can be distinguished in the corpus. These patterns correspond to four different degrees of structural complexity of the IESCx, and take into account the presence or absence of different kinds of modifiers and arguments of the verbal nucleus *sabes*.[15]

The two ends of the structural complexity continuum are occupied by the bare pragmatic marker *sabes* and the matrix clause Cx respectively. On the one hand, the lowest degree of formal complexity (indicated as *zero*) is occupied by *sabes* without any modifiers. On the other hand, the highest degree can be attributed to those constructional variants in which there is a relationship of subordination between the expression of stance by *sabes* in the main clause and the host introduced by a complementizer (viz. the conjunction *que* 'that' or the relative *lo que* 'what' as in *tú sabes que* + clause, *como sabes que* + clause, *ya sabes tú que* + clause etc.).

In between, we find a wide gamut of parenthetical constructions (*tú sabes, lo sabes, sabes una cosa, como sabes,* etc.), that are syntactically detached from their host, and cause syntactic and/or prosodic disruption.[16] This corresponds to the general definition of parenthetical clauses which can be

[15] Note that this classification is just one of several ways to define the various degrees of complexity of the IESCx. A more fine-grained classification identifies ten degrees of complexity, taking into account not only the argument status and formal length of the modifiers but also the presence of one, two or more of them. However, this classification proved to be too detailed for the present purposes as it disguised more general tendencies in the data.

[16] The exact relationship between matrix clauses and parentheticals has been disputed in the literature. According to the "matrix clause hypothesis" (Thompson and Mulac 1991), parentheticals derive from subordinating matrix clauses through a process of syntactic reanalysis. However, this view has been disputed by Brinton (2008) and Van Bogaert (2009) who, based on diachronic empirical data, argue in favor of a more complex picture, including adverbial clauses and imperative clauses as origins. For the Spanish cognates, the question also remains unanswered but will be addressed in more detail in future studies.

summarized as follows (e.g. Thompson and Mulac 1991; Schneider 2007, 2014; Brinton 2008; Van Bogaert 2009; Kaltenböck 2013):
1. parentheticals lack a syntactic connection with their host, to which they are related only by adjacency; as a consequence, they have a very mobile position;
2. they are semantically independent from the host, which implies that they can be added or left out without influencing the grammaticality of the clause;
3. prosodically, they are often characterized by a "comma intonation" implying a lower pitch, preceding and following pauses and a rising terminal. Note, however, that not all parentheticals are preceded or followed by pauses as they can also be prosodically integrated in their host, with which they adopt an "enclosing" melody (see Dehé 2007: 270–274 on the epistemic parentheticals *I think/I believe*; Avanzi 2012; Schneider 2014: 280-281)

However, with regard to their structural complexity, parenthetical Cxs can be placed in two different categories. In a first class, the nucleus *sabes* is modified by the negative adverb *no*, the relational adverb *como,* or the time adverb *ya*. Given that these modifiers have a rather peripheral status with regard to the argument structure of the verbal nucleus, these instances are classified as representing a *low* degree of formal complexity of the construction. The second class consists of those parentheticals where the verbal nucleus is accompanied by one, or both, of its arguments, with (*ya sabes tú, ya lo sabes, tú no sabes* etc.) or without (*tú sabes, lo sabes, sabes una cosa, sabes qué* etc.) other adverbial modifiers, when it is defined as displaying a *medium* degree of complexity.

To recap, the large list of IES expressions (provided in Section 3.1), can be sorted into four subschemas, corresponding to various degrees of structural complexity:
1. matrix clause Cxs (high degree of formal complexity): *sabes lo que V, sabes que + clause, tú sabes lo que V, sabes tú lo que V, tú sabes que + clause, sabes tú + clause, ya sabes que + clause, no sabes lo que V, ya sabes lo que V,…*
2. parenthetical Cxs with expression of arguments (medium degree) : *tú sabes, sabes tú, como tú sabes, como sabes tú, sabes qué, no lo sabes, ya lo sabes, tú no sabes, lo sabes, sabes (una) cosa,…*
3. parenthetical Cxs with adverbial modifiers (low degree): *ya sabes, sabes ya, no sabes, como sabes, como ya sabes,…*
4. frozen pragmatic marker *sabes* (zero).

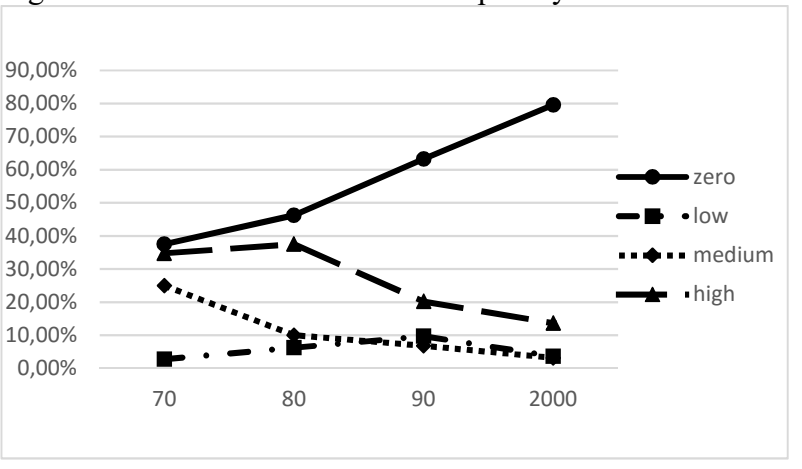
Table 3 and Figure 3 detail the formal variation of the IESCx throughout the four diachronically defined subcorpora.[17]

[17] As has been suggested by an anonymous reviewer, the categories of medium and low complexity could be combined, as it would result in a conceptually and quantitatively sharper division. However, this three-part classification would prevent us from comparing the evolution of parentheticals including arguments vs. adverbial modifiers, a distinction which will be shown to be relevant for the purpose of this article.

Table 3. Evolution of structural complexity of IESCx

	high (MClCx)		medium (PCx)		low (PCx)		zero (*sabes*)		total	
	#	%	#	%	#	%	#	%	#	%
70s	25	34,72%	18	25,00%	2	2,78%	27	37,50%	72	100
80s	30	37,50%	8	10,00%	5	6,25%	37	46,25%	80	100
90s	206	20,18%	70	6,86%	99	9,70%	646	63,27%	1021	100
>2000	287	13,63%	67	3,18%	76	3,61%	1675	79,57%	2105	100

Figure 3. Evolution of structural complexity of IESCx



The data tell us that the evolution of the IESCx is characterized by a gradual erosion of the modifiers of the verbal nucleus. In general terms, the earlier decades (70s-80s) display a balanced frequency distribution between high and zero complexity structural variants. From the 1990s on, the latter category has increasingly become the dominant one. More specifically, in the 70s corpus, despite its small size, the two most divergent categories are well represented with 37.50% of all tokens of the construction being of zero complexity and 34.72% of them falling into the highest complexity category. The medium parenthetical Cx is also rather frequently attested (25%), while the low complexity category is virtually absent (2.78%). In the 80s corpus, the medium parenthetical Cx appears to have been sharply reduced (10%), while the frequency of the low (6.26%) and zero complex categories (46.25%) have risen. This evolution continues in the 90s corpus, which, however, also shows a notable decrease in the high complexity type (20.18%). Finally, in the >2000 corpus all categories have shrunk for the benefit of the zero complexity class, the frozen *sabes* (79.57%).

There has thus been an important change in the relative frequencies of the constructional variants of the IESCx over the last 50 years. The quantitative data suggest that the construction as a whole has undergone a process of structural reduction. Moreover, this reduction of the IESCx seems to have followed a cascade-like model. Initially, the erosion process seems to have left aside the adverbial modifiers and to have mainly affected the arguments. Moreover, it set off with the omission of (pro)nominal arguments (e.g. *lo* 'it',

una cosa 'something'*, tú* 'you'). Indeed, the medium parenthetical Cx is the only construction variant that has consistently become less frequent over the four time periods; the matrix clause Cx has put up stronger resistance in the earliest decades (70s and 80s) before also rapidly losing ground. The class of low complexity variants, including adverbial modifiers, seems to spread until the 90s, but shrinks in present-day Spanish.

This study thus reveals that the diachronic evolution and formal changes within a more generally defined constructional schema do not come about arbitrarily; rather, they follow a morphologically and syntactically motivated order. The four different categories are still attested in present day Spanish, but it is to be expected that the zero category will gain in strength in the following years and decades.

As a consequence, the configuration of the constructional network itself seems to have evolved from a well-balanced distribution between formally different variants in the 70s, towards a configuration in which one variant, namely the bare *sabes*, has become the most salient one in present-day Spanish. The formally complex structures have gradually moved towards the periphery of the network.

However, it is not enough to provide an answer to the question whether or not and in what directions the configuration of the IESCx have changed. The - perhaps even more interesting - question of *why* these changes have occurred should also be addressed.

5.2 Accounting for the reconfiguration of the network

The proposal by Haiman (1994), in which he argues for a parallelism between general cultural phenomena of ritualization and the process of grammaticalization, gives some pretty good clues as to how to deal with this issue. The central idea that the frequent repetition of a certain practice entails "habituation" and "emancipation" of that practice is particularly inspiring for understanding the structural changes to which the IESCx has been subject. It shows that, in fact, these relate to more general functional changes within the domain.

Indeed, just as the frequent repetition of a cultural practice leads towards a depletion of its original force and significance (e.g. making the sign of a cross now has more than its literal religious meaning), so the speech act through which, originally, a speaker asks the interlocutor to verify his knowledge on some upcoming or preceding information becomes less meaningful (and thus a "habit") when frequently repeated. Consequently, parallel to frequently repeated practices, the expressions concretizing this speech act (such as *sabes lo que pasa*, *no sabes, como tú sabes* etc.) gain a more symbolic ("emancipated") function, and lose (a part of) their original "instrumental" (interpersonal and epistemic) function. To put it differently, the expression of stance moves from the foreground to the background in the flow of information.

As argued by Bybee (2003: 605), repetition is crucial for understanding universal grammaticalization paths. The same mechanism seems to (partially) explain internal reconfigurations within the network of a more general schema. She goes further:

> the explanation for the content of what is repeated requires reference to the kinds of things human beings talk about and the way they choose to structure their communications. (Bybee 2003: 605)

From this viewpoint, the answer to the question of why the interlocutor-oriented questions, to wit, the constructs of the IESCx, have been – and still are – so entrenched in language is straightforward: it comes down to recognizing the inherently interpersonal nature of human communication.

The IESCx being a high (and increasing) frequency construction, it is now easy to understand why it has been affected by a process of internal reduction. As the meaning of the construction weakens, the informational contribution of its constituent elements also decreases. In the first place, functionally "redundant" elements such as the expletive 2nd person pronoun *tú* (in Spanish morphologically indicated in the verb suffix), or semantically empty direct objects, such as *lo* 'it'*, una cosa* 'a thing'*,* or *qué* 'what', are omitted. In a second phase this reduction process also extends towards the subordinate clause in the matrix clause Cx, which, often, is also semantically poor (e.g. *sabes lo que pasa* 'do you know what happens', *sabes lo que pienso* 'do you know what I think'). The adverbial modifiers, although not that frequent in the corpus, seem to be more resistant to this process of formal reduction. As a hypothesis it can be stated that this is due to their overt *framing* function, namely, the fact that they explicitly refer to the cognitive status of the interlocutor: previous knowledge in *como sabes* 'as you already know' and *ya sabes* 'you already know', doubtful knowledge in *¿no sabes?* 'don't you know?'. In a later phase, they are also subject to erosion.

In sum, once entrenchment of the interpersonal epistemic stance schema occurred, its frequency of use has had an effect on its form. The repeated variants of the construction have become more integrated and reduced, and this process has followed a morpho-syntactically (shorter (pro)nominal items > subordinate longer items), and semantically conditioned (semantically poor element > framing elements) order.

6. Conclusion

Summing up, the data analyzed in this study provide valuable insights concerning the three goals that have been set out in its introduction: (1) to offer a theoretical contribution to the studied discourse related phenomena from a constructional perspective, (2) to provide a clear overview of the taxonomy of interpersonal epistemic expressions and (3) to give a

methodological illustration of how recent language change can be studied in Spanish.

First, it has been shown that the constructional perspective can be successfully applied to the domain of pragmatic markers and related expressions. A lexical item such as *sabes* is not to be analyzed as if it were an isolated sequence, but rather as part of a wider taxonomy. This approach motivates a reformulation of the more basic question whether or not a particular lexical item has reached the status of pragmatic marker. The questions become "how has the schema or network to which the item pertains developed over time?" and also "Can it be asserted that the lexical form has gradually distinguished itself from its family members?" This article has thus presented a multidimensional approach to the study of the Interpersonal Epistemic Stance Construction in Spanish, meaning that it has focused not only on the development of particular constructs at a lower level, but also on the evolution of the schema at a higher level. It has clearly shown that the grammaticalization account of pragmatic markers and a diachronic network approach go hand in hand.

Second, the case study itself was based on detailed observations of the frequency of use of both the schema and its constructional variants. This quantitative study has afforded a more fine-grained approach to the relationship between frequency and entrenchment, and has provided a (partial) window into the Spanish speaker's linguistic organization of the interpersonal epistemic speech act. Once again, it has been confirmed that frequency effects can explain particular phenomena of language change. The extending use of interpersonal expressions over the last decades has led towards a "routinization" of a sequence of words, which in its turn has led to an increased use of more formally reduced or eroded forms.

This study has also shown that variability, and thus competition between variants, within a more general schema is important for understanding (recent) change. The analysis of whether the sequential decades yield similar or different frequencies of use of these variants has shown that the "frozen *sabes*" form has gradually invaded the territory of the more formally complex alternatives. Over the last 50 years, there has thus been a reform of the internal configuration of the network: the most eroded form has moved towards the center, pushing aside more formally complex expressions. However, this study has shown that the process has not occurred at random but has followed a morpho-syntactically and semantic-pragmatically motivated order.

Third, it follows from this case study that the end of the 20th and the beginning of the 21st century can rightly be studied as a locus of current language change in Spanish. It is important to note that these recent language changes can been correlated with changes in the "overarching sociohistorical context":

> Geographic and economic mobility, shifting social norms as well as the current revolution in communication technology all contribute to the

> acceleration of certain types of linguistic change at the turn of the twenty-first century. (Aarts et al. 2013:6)

Spanish society has indeed lived through a number of important economic, social and cultural revolutions which started during the second half of the 20th century. These include: the development of an industrial urban society, the advent of mass media, an increase in general educational level which resulted in sociocultural level becoming an important linguistic variable, and, last but not least, the steady process of "colloquialization" of Spanish (analogous to that in other languages such as English, see Aarts et al. 2013). This last feature involves a decline in the use of the high register and a levelling effect of the informal register (Koch and Oesterreicher 2007: 234-235; Pons Bordería 2014: 1000), in both spoken and in written language. As a consequence, the 20th century and the beginning of the 21st century should not be overlooked as a site of linguistic change. It has already been observed that, in this period, the development of discourse markers in Spanish has considerably accelerated (e.g. the pragmaticalization of *o sea* 'that is to say', *entonces* 'so', *encima* 'what is more', *vaya* or *ándale* 'come on', e.g. Pons Bordería 2014). The case of *sabes* further corroborates this observation.

References

Aarts, Bas, Joanne Close, Geoffrey N. Leech & Sean Wallis (eds). 2013. *The Verb Phrase in English. Investigating Recent Language Change with Corpora*. Cambridge: University Press.

Aijmer, Karin. 1997. I think – an English Modal Particle. In Toril Swan & Olaf Jansen Westvik, eds. *Modality in Germanic Languages: Historical and Comparative Perspectives.* Berlin, New York: Mouton de Gruyter, 1-47.

Avanzi, Mathieu. 2012. La prosodie des verbes parenthétiques en français parlé. In Julie Glikman & Mathieu Avanzi, eds. *Entre rection et incidence: des constructions verbales atypiques? Études sur* je crois*,* je pense *et autres parenthétiques*. Nanterre: Université Paris Ouest Nanterre La Défense, 131-144.

Beeching, Kate & Ulrich Detges. 2014. Introduction. In Kate Beeching & Ulrich Detges, eds. *Discourse functions at the left and right periphery: crosslinguistic investigations of language use and language change.* Leiden: Brill, 1-23.

Behrens, Heike & Stefan Pfänder. 2016. *Experience Counts: Frequency Effects in Language*. Berlin & Boston: Mouton de Gruyter.

Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan (eds). 1999. *Longman Grammar of Spoken and Written English.* Harlow & Essex: Pearson Educational.

Brinton, Laurel J. 1996. *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin & New York: Mouton de Gruyter.

Brinton Laurel J. 2008. *The Comment Clause in English. Syntactic Origins and Pragmatic Development.* Cambridge: Cambridge University Press.

Briz Gómez, Antonio & Salvador Pons Bordería. 2010. Unidades, marcadores discursivos y posición. In Óscar Loureda & Esperanza Acín, eds. *Los estudios de marcadores del discurso en español, hoy.* Madrid: Arco Libros, 327-358.

Bybee, Joan. 2003. Mechanisms of Change in Grammaticization: The Role of Frequency. In Brian D. Joseph & Richard D. Janda, eds. *The Handbook of Historical Linguistics*Oxford: Blackwell Publishing Ltd, 602-623.

Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Cameron, Richard. 2011. "Aging, Age, and Sociolinguistics". In Manuel Díaz-Campos, ed. *The Handbook of Hispanic Sociolinguistics*? West Sussex: Blackwell, 207-229.

Cappelle, Bert. 2006. Particle Placement and the Case for 'Allostructions'. *Constructions* 1: 1-28.

Company Company, Concepción. 2006. Subjectification of Verbs into Discourse Markers: Semantic-Pragmatic Change Only? In Bert Cornillie & Nicole Delbecque, eds. *Topics in Subjectification and Modalization*. Amsterdam: John Benjamins, 97-121

Cuenca, Maria Josep. 2013. The Fuzzy Boundaries between Discourse Marking and Modal Marking. In Liesbeth Degand, Bert Cornillie & Paola Pietrandrea, eds. *Discourse Markers and Modal Particles. Categorization and Description*. Amsterdam: John Benjamins, 191-216.

Degand, Liesbeth & Benjamin Fagard. 2011. *Alors* between discourse and grammar: The Role of Syntactic Positon. *Functions of Language* 18 (1): 29-56.

Dehé, Nicole. 2007. The relation between syntactic and prosodic parenthesis. In Nicole Dehé & Yordanka Kavalova, eds. *Parentheticals*. Amsterdam: John Benjamins, 261-284.

Dehé, Nicole & Anne Wichman 2010. The Multifunctionality of Epistemic Parentheticals in Discourse. Prosodic Cues to the Semantic-pragmatic Boundary. *Functions of Language* 17 (1): 1-28.

Erman, Britt. 2001. Pragmatic markers revisited with a focus on *you know* in adult and adolescent talk. *Journal of Pragmatics* 33: 1337-1359.

Dostie, Gaétane & Suzanne de Sève. 1999. Du savoir à la collaboration. Étude pragma-sémantique et traitement lexicographique de *t'sais*. *Revue de Sémantique et Pragmatique* 5: 11-35.

Enghels, Renata & Marlies Jansegers. 2013. On the cross-linguistic equivalence of *sentir(e)* in Romance languages: a contrastive study in semantics. *Linguistics* 51 (5): 957-991.

Enghels, Renata, Clara Vanderschueren & Miriam Bouzouita. 2015. Panorama de los corpus y textos del español peninsular contemporáneo.

In Eugeen Roegiest & Maria Iliescu, eds. *Manuel des anthologies, corpus et textes romans* 7, Berlin & New York: Mouton de Gruyter, 147-170.
Fischer, Kerstin. 2015. Conversation, Construction Grammar, and Cognition. *Language and Cognition* 7: 563-588.
Fraser, Bruce. 1999. What are discourse markers? *Journal of Pragmatics* 31: 931-952.
Fried, Mirjam & Jan-Ola Östman. 2005. Construction Grammar and Spoken Language: the Case of Pragmatic Particles. *Journal of Pragmatics* 37 (11): 1752-1778.
Goldberg, Adele 1995. *Constructions: a Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
Geeraerts, Dirk. 2010. The doctor and the semantician. In Glynn, Dylan & Kerstin Fischer, eds. *Quantitative methods in cognitive semantics: Corpus-driven* approaches. Berlin, New York: Mouton De Gruyter, 63-78.
Glynn, Dylan & Kerstin Fischer (eds.). 2010. *Quantitative methods in cognitive semantics: Corpus-driven approaches.* Berlin & New York: Mouton De Gruyter.
Gries, Stefan Thomas & Dagmar Divjak (eds.). 2012. *Frequency Effects in Language Representation*. Berlin, Boston: Mouton De Gruyter.
Haiman, John. 1994. Ritualization and the Development of Language. In William Pagliuca, eds. *Perspectives on Grammaticalization.* Amsterdam: John Benjamins, 3-28.
Hernández, José Esteban & Rubén Baldazo. 2013. Distribución variable de un marcador del discurso: No decir nada es como decir *verdad*. *Revista Española de Lingüística* 48 (2): 113-150.
Hilpert, Martin. 2013. *Constructional Change in English. Developments in Allomorphy, Word Formation, and Syntax.* Cambridge: Cambridge University Press.
Himmelmann, Nikolaus, 2004. Lexicalization and Grammaticalization: Opposite or Orthogonal?. In Walter Bisang, Nikolaus Himmelmann & Björn Wiemer, eds. *What Makes Grammaticalization. A Look from its Fringes and Its Components*. Berlin: Mouton de Gruyter, 19-40.
Hoffmann, Thomas & Graeme Trousdale. 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.
Hopper, Paul J. 1991. On Some Principles of Grammaticization. In Elizabeth C. Traugott & Bernd Heine, eds. *Approaches to Grammaticalization.* Amsterdam: John Benjamins, 17-36.
Hopper, Paul J. & Elizabeth C. Traugott. 2003. *Grammaticalization.* Cambridge: Cambridge University Press.
House, Juliane. 2009. Subjectivity in English as Lingua Franca Discourse: The Case of *you know*. *Intercultural Pragmatics* 6 (2): 171-193.
Jucker, Andreas H. 1995. *Historical Pragmatics*. Amsterdam: John Benjamins.
Kaltenböck, Gunther. 2013. The Development of Comment Clauses. In Bas Aarts, Joanne Close, Geoffrey N. Leech & Sean Wallis, eds. *The Verb*

Phrase in English. Investigating Recent Language Change with Corpora. Cambridge: Cambridge University Press, 286-317.
Kärkkäinen, Elisa. 2003. *Epistemic Stance in English Conversation: A Description of Its Interactional Functions, with a Focus on I Think.* Amsterdam: John Benjamins.
Koch, Peter & Wulf Oesterreicher. 2007. *Lengua hablada en la Romania: español, francés, italiano*. Madrid: Gredos.
Lindström, Jan & Camilla Wide. 2005. Tracing the Origins of a Set of Discourse Particles: Swedish Particles of the Type *you know*. *Journal of Historical Pragmatics* 6: 211– 236.
López Serena, Araceli & Margarita Borreguero Zuloaga. 2010. Los marcadores del discurso y la variación lengua hablada vs. lengua escrita. In Óscar Loureda Lamas & Esperanza Acín Villa, eds. *Los estudios sobre marcadores del discurso en español, hoy*. Madrid: Editorial Arco Libros, 415-495.
Martín Zorraquino, María Antonia & José Portolés Lázaro. 1999. Los marcadores del discurso. In Ignacio Bosque, and Violeta Demonte, eds. *Gramática descriptiva de la lengua española*. Madrid: Espasa Calpe, 4051-4214.
Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical Sociolinguistics: Language Change in Tudor and Stuart England.* London: Pearson Education.
Nuyts, Jan. 2001. *Epistemic Modality, Language and Conceptualization: a Cognitive Pragmatic Perspective*. Amsterdam: John Benjamins.
Pons Bordería, Salvador. 2014. El siglo XX como diacronía: intuición y comprobación en el caso de *o sea*. *RILCE* 30 (3): 85-1016.
Schneider, Stefan. 2007. *Reduced Parenthetical Clauses as Mitigators. A Corpus Study of Spoken French, Italian and Spanish*. Amsterdam: John Benjamins.
Schneider, Stefan. 2014. Parenthesis: fundamental features, meanings, discourse functions and ellipsis. In Marlies Kluck, Dennis Ott & Mark de Vries, eds. *Parenthesis and ellipsis: crosslinguistic and theoretical perspectives*. Berlin: De Gruyer Mouton, 277-300.
Schwenter, Scott A. 1996. Some reflections on *o sea*: A discourse marker in Spanish. *Journal of Pragmatics* 25: 855-874.
Tanghe, Sanne. 2016a. Position and polyfunctionality of discourse markers: the case of Spanish markers derived from motion verbs. *Journal of Pragmatics* 93: 16-31.
Tanghe, Sanne. 2016b. *Marcadores derivados de verbos de movimiento. Una aproximación cognitiva a su polifuncionalidad*. Berlin: Mouton de Gruyter.
Thompson, Sandra A. & Anthony Mulac. 1991. The Discourse Conditions for the Use of the Complementizer *that* in Conversational English. *Journal of pragmatics* 15: 237-251.

Torres Cacoullos, Rena & Scott A. Schwenter. 2007. Towards an operational notion of subjectification. *Proceedings of the 31st Annual Meeting of the Berkeley Linguistics Society*, 347-358.

Torres Cacoullos, Rena & James A. Walker. 2011. Collocations in Grammaticalization and Variation. In Bernd Heine & Heiko Narrog, eds. *Handbook of Grammaticalization.* Oxford: Oxford University Press, 225-238.

Traugott, Elizabeth C. 2007. Discussion Article: Discourse markers, Modal Particles, and Contrastive Analysis, Synchronic and Diachronic. *Catalan Journal of Linguistics* 6: 139-157.

Traugott, Elizabeth C. 2014. On the Function of the Epistemic Adverbs *Surely* and *no doubt* at the Left and Right Peripheries of the Clause. In Kate Beeching & Ulrich Detges, eds. *Discourse functions at the Left and Right Periphery. Crosslinguistic investigations of language use and language change*. Leiden, Boston: Brill, 72-91.

Traugott, Elizabeth C. In press. Modeling Language Change with Constructional Networks.

Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and Constructional Change*. Oxford: Oxford University Press.

Trousdale, Graeme. 2014. On the relationship between grammaticalization and constructionalization. *Folia Linguistica* 48 (2): 557–577.

Urzúa-Carmona, Paula. 2006. El Verbo 'cachar' en el español coloquial de Chile. *Onomázein* 13 (1): 97-107.

Van Bogaert, Julie. 2009. *The Grammar of Complement-taking Mental Predicate Constructions in Present-day Spoken British English.* PhD dissertation, Ghent University.

Van Bogaert, Julie. 2010. A Constructional Taxonomy of *I think* and Related Expressions: Accounting for the Variability of Complement-taking Mental Predicates. *English Language and Linguistics* 14 (3): 399-427.

Vázquez Rozas, Victoria. 2006. Construcción gramatical y valor epistémico. El caso de *supongo*. In Milka Villayandre Llamazares, ed. *Actas del XXXV simposio internacional de la sociedad española de lingüística.* León: Universidad de León, 1888-1900.

Zeldes, Amir 2012. *Productivity in Argument Selection. From Morphology to Syntax*. Berlin: Mouton de Gruyter.

Zimmerman, Klaus. 2002. La variedad juvenil y la interacción verbal entre jóvenes. In Félix Rodríguez González, ed. *El lenguaje de los jóvenes*. Barcelona: Ariel, 17-26.