

Determining prefab status in New Mexican Spanish, the case of *decir*

Víctor A. Valdivia-Ruiz
University of New Mexico

ABSTRACT. The present study addresses the issue of how to determine what a prefab is; specifically, it proposes features such as frequency, morphosyntactic variability, and prosodic distribution, constitute operational measures with which structural gradience can be accounted for. Results from the analysis of *decir* clauses from the *New Mexico and Colorado Spanish Survey* (Bills & Vigil 2008) show subject expression and position, verb's position, and prosodic distribution between the verb and its complement are important indicators that a particular subject-tense collocation is a prefab.

Keywords: prefab, variability, intonation unit

1. INTRODUCTION. Prefabs are multiword sequences, conventionalized expressions 'known to speakers as such' (Bybee and Beckner 2009:835), but whose meaning may or may not be idiosyncratic. These complex units are of particular interest within functional-cognitive approaches to linguistics, especially within the USAGE-BASED THEORY, not only because they make evident the need to develop a theory that deals with pairings of form-meaning larger than the word, but also because of their relevance in the general-domain of analogy. Since speakers' experiences affect both categorization of linguistic input and production of new utterances (Bybee 2010:8), prefabs provide a large repertoire of elements for comparison given the vast and complex networks they can be part of by establishing links with other instances both through their conventionalized meaning as a whole and through their internal lexical components.

As with all other constructions, prefabs result from chunking, a process by which the repetition of two or more words together causes them to be interpreted and later produced, as a whole unit. The central role of repetition implies the process is gradual, so linguistic units move along a continuum from one category to another, and it is not always easy to distinguish the boundaries of many categories of grammar. In the case of prefabs, gradience allows us to include in the category semantically transparent expressions such as 'black and white' but also others as 'to rub the wrong way', which cannot be understood by simply adding the individual meanings of each one of its parts. Because prefabs are gradually developing units, the emergence of a new one does not mean the old one disappears; they coexist as long as they both continue being used in discourse, and the boundaries between both of them are blurred. Furthermore, it is often the case that some of the lexical meanings of one or more items in the prefab are still observable, which allows them to establish links with other constructions containing semantically similar items (Bybee & Torres 2009).

Furthermore, constructions may 'include not only properties of the situation described by the utterance but also properties of the discourse in which the utterance is found [...] and of the pragmatic situation of the interlocutors' (Croft 2001:14). Thus, the more a prefab is used in particular communicative situations, the more semantic bleaching it experiences, and, consequently, the more likely it will convey pragmatic and even epistemic nuances.

Based on the former characteristics and phenomena, I will argue there is evidence suggesting some multi-word sequences involving *decir* are developing prefab status, and that the former can be determined by a quantitative analysis of various morphosyntactic and prosodic features such as subject expression and position; complement type, realization and position; co-

occurrence of the verb and its arguments in the same intonation unit; and complement's distribution in one or more intonation unit, among others. Criteria for the operationalization of these features will be discussed in detail in section 4.

2. DATA. Data for this paper were extracted from 18 one-on-one interviews from Bills & Vigil's 2008 *New Mexico and Colorado Spanish Survey* (NMCROSS). The duration of speakers' formal instruction varies from minimal schooling to 17 years, and their professions are very diverse. Data from interviewers have also been included in the study because, as members of the community to be studied, they provide valuable data as their speech may show discourse functions only observable in their interaction with the interviewees' speech; for instance, the common use of questions and their goal of eliciting speech from the interviewee. It is important to mention that code-switching is not unusual in the interviews.

Interviews were transcribed according to Du Bois and colleagues' 1993 system, in which each line represents an intonation unit (IU), a speech segment uttered under one intonation contour.¹ Even though IU's do not directly correspond to syntactic units such as phrases or clauses, they do appear to play a role in the way grammar is organized (Chafe 1994, Ono & Thompson 1995) and thus are incorporated into the analysis. Using the software MonoConc (Barlow 1999), data were extracted from approximately 8 hours, 10 minutes of recording, corresponding to 87,800 words and 24,544 IU's.

3. METHODOLOGY. The sample was extracted following two criteria. The first one was to exclude clauses in which the verb or its arguments were unclear, or incomplete.² Both phenomena are illustrated by 1 below:

- (1) Cuando se fueron,
... (2.0) dijo,
<X hallo el X>--
'When they left, she said, <X I find the X> --'³

The second criterion was to have a similar amount of tokens from each interview. In order to reach 300 tokens, approximately 17 from each of the 18 interviews were needed. However, some interviews did not have 17 tokens, and thus, more than 17 tokens were taken from others. The minimum extracted from one interview was six, and the maximum 19 tokens. For interviews with more than 17 tokens available, the first 17 clauses were extracted.

Each clause was coded for a series morphosyntactic and prosodic features (see section 4) in order to identify both the most frequent collocations and their structural variability. The latter is particularly important in cases in which two or more collocations are constructed around the same verbal form since it provide clues regarding whether they all represent the same prefab, or whether they are indeed different ones.

4. CODING. For the analysis, the basic units are the clause and clause combinations, understood as the syntactic structure formed by a verb and its arguments. This follows previous research on structural patterns (e.g. Scheibman 2001, 2002; Travis 2006), which will allow a clearer and more precise comparison between those results and the ones from this study. In order to identify what factors affect morphosyntactic variability, as well as the role of cognitive management of information, each clause was coded for the following features:

- a. Subject person and number (1SG, 2SG, 3SG, 1PL, etc.)⁴
- b. Subject realization (pronouns, full nominal phrases, etc.)
- c. Subject position (SV or VS)
- d. Verbal structure (simple, modal, and periphrastic clauses)
- e. Tense, aspect, mood (TAM)
- f. Type of complement⁵
- g. Complement's position (VC, CV, CVC)
- h. Verb and complement's prosodic distribution⁶
- i. Complement's prosodic distribution⁷
- j. Indirect object (occurrence vs. non-occurrence, morphological realization)

5. RESULTS. Three verbal tenses account for most of the clauses in the data: present (122/300, 40.6%), preterit (95/300, 31.66%) and imperfect (66/300, 22%); however, only the first group gave rise to collocations developing prefab status, even if their token frequency is not particularly high.⁸ These collocations are: *(yo) digo* 'I say', *digo yo* 'I say, lit. Say-I', *como (LE) digo* 'as I say / as I told you / I'm telling you', and *le dicen X* 'they call it X'.

These results do not mean the other groups did not produce collocations with certain discourse functions, but they constitute more schematic (i.e. general) patterns. For instance, although *(él/ella) dijo* 'he/she said' is the most frequent item in the data (54/300: 18%), as well as the central exemplar for both preterit (54/95: 56.84%), and 3SG clauses (54/92: 58.69%), it shows a high degree of morphosyntactic variability; hence, there is not a specific collocation strong enough to cause structural fixation.⁹

5.1. (YO) DIGO 'I SAY'. This item represents 53% (26/49) of the 1SG-present group. It includes both unexpressed (21/49: 43%) and expressed subjects (5/49: 10%) in preverbal position. The preponderance of *(yo) digo* is observed too in the oral section of the database *Corpus del Español* (Davies 2002), where it is the second most frequent form for *decir* in present tense: *digo* appears 2,432 times, and *yo digo* 422 times. Together, these forms represent 32.5% (2854/8784) of the group.

In most cases, the collocation has a direct object (22/26: 85%), usually realized by a clause, or a sentence (19/26: 73%). These finite complements tend to appear (14/19: 74%) in the canonical postverbal position. Regarding indirect objects, they appear when the speaker uses *(yo) digo* 'I say' to refer to an utterance from a previous speech event in which there was a specific addressee:

- (2) Yo **les** digo,
Yo no nací pa' sentada. (219-1A2: 224)
'I tell **them**, I wasn't born [to be] sitting.'

When the indirect object is not included, the utterance is not addressed to anyone in particular, not even the interviewer, because the speaker is expressing a belief, opinion, or decision, as in 3 and 4; thus, *decir* 'to say' does not reproduce a speech event, but a thought. Given the variation of meaning associated with the presence or absence of an indirect object pronoun, it is plausible to think two different prefabs are emerging from the collocation *(yo) digo* 'I say'. More data are necessary to confirm the hypothesis.

- (3) yo también digo que **gavilán es hawk**. (270-1B2: 93)
 ‘I too say that *gavilán* means hawk’
- (4) B: What advice would you [leave them]?
 W: [Pues yo],
 yo digo=,
 .. que yo,
todo lo que tengo,
yo --
yo me ha tenido --
yo ha tenido que trabajar. (190-3B2: 19)
 B: ‘What advice would you leave them?’
 W: ‘Well, I, I say that I, **everything I've, I, I've had, I've had to work** [for it].’

A feature that in this data is shown to be central to determining whether a collocation is developing prefab status is the prosodic distribution between the verb and its complement. More than half of the complements for (*yo*) *digo* ‘I say’ (15/26: 58%) appear in a different intonation unit than the verb, and most of them are finite complements (14/19: 74%), as in 5. If (*yo*) *digo* ‘I say’, being a low content verb, does not convey new information, then what is the cause for such prosodic independence?

- (5) Yo les digo,
Yo no nací pa' sentada. (219-1A2:224)
 ‘I tell them, **I wasn't born [to be] sitting.**’

For several of the tokens, the explanation lies in the use of the prefab as a direct speech device. This device attempts to eliminate the cognitive effects associated with the displaced mode in the discourse, the mode in which the speaker focuses on experiences that originated from past events (Chafe 1994:199). Moving from this mode to the immediate mode means the speaker pretends that the experience is perceived either as occurring at the time of its representation, or as relevant to it. The effect is stronger when direct speech is used together with historical present, as in 6.

Furthermore, the immediate move is where the speaker can access, evaluate, and regulate the idea being communicated, instead of merely remembering or imagining, as in 5 above. In the interview, the speaker is talking about how her kids want her to stop weaving and farming because of her age. Thus, when she says *yo no nací pa' sentada* ‘I wasn't born to be sitting’, she is offering an explanation for her refusal to follow her kids’ wishes. Additionally, the utterance does not refer to a single previous speech event, but rather to an idea expressed habitually. The important thing here is not whether the speaker says exactly the same thing time after time, but the repetition of an idea. Thus, when using direct speech, the speaker is performing two processes at the same time, one with the verb and the other with the complement. Since each of them has its own cognitive load, they tend not to share an intonation unit.

In the data, 63% (12/19) of all (*yo*) *digo* ‘I say’ clauses with finite complements show this prosodic structure, suggesting such subject-verb collocations often perform one of the functions described above. Moreover, since the event is presented as if occurring at the time of its expression, it exists by itself, therefore it is not syntactically dependent on (*yo*) *digo* ‘I say’. Such

a loss of dependence allows cases in which the verb appears inside the complement, or in which two instances of *digo* ‘I say’ refer to the same complement, as in 6 and 7, respectively.

(6) *quítate el cute y,*

.. No,

le digo,

ya me voy pa' mi casa. (318-1A3: 160)

‘[He said] Take off your coat. *No, I tell him, I’m going home. ’*

(7) (H) *Y luego cuando acabó con mi hermana le digo,*

pues,

eh,

no me puede pegar a mí porque yo no estoy con usted, le digo. (318-1A3: 143)

‘And then, when he was done beating my sister, *I tell him: “well, you have no right to hit me because I do not live with you”, I tell him. ’*

Although few in number, utterances such as 6 and 7 are very relevant for the argument about the prefab status of (*yo*) *digo* for two reasons. Firstly, the fact that the collocation is intercalated into the complement indicates loss of syntactic cohesion between both constituents; that is to say, they seem to be structurally independent from each other rather than two elements of a sentence. Secondly, structural changes have semantic and functional effects. Thus, using (*yo*) *digo* ‘I say’ when the information was already or is being presented implies the collocation is doing more than merely reporting information. For instance, in 6 and above, by recreating the actual wording of the original discussion, the speaker also indicates his involvement and stance (cf. Chafe 1994). Moreover, in intra-clausal position, the collocation can convey epistemicity without having a textual function, as in 8, where *yo digo* has a meaning closer to ‘I mean’ than to ‘I say.’

(8) A: *Yo también lo <X digo para decir X>,*

(H) *porque nos pasamos tanto tiempo en el teléfono,*

digo,

Y: *Sí.*

A: *Y luego no caminamos, (147-2A2: 133)*

A: ‘Me too, <X I said just as an excuse X>, because we spent a lot of time on the telephone, *I mean, ’*

Y: ‘Yes’.

A: ‘ *and then we don't go for our walk... ’*

In a previous part of the conversation, speaker A is talking about one of her friends who recently bought a cordless telephone. Then, she states she wants one too (first line) and offers the reason for the potential purchase (second line) followed by *digo* ‘I mean’ (lit. ‘I say’). Finally, after an interruption by the other person, Alicia mentions the consequences of spending too much time on the telephone (fifth line). The complexity of this segment of conversation clearly illustrates that of (*yo*) *digo* ‘I say’. First, it is not clear whether the collocation refers to the clause in line 2 or to the one in line 5, or whether both clauses compose a whole explanation. Second, although the situation expressed by Alicia (i.e. spending too much time on the telephone) is

something habitual and relevant to the present, it really does not refer to a specific speech event, so *(yo) digo* does not function as direct or indirect speech. Third, even if we consider both lines 2 and 5 as parts of one explanation, the intercalation of *(yo) digo* suggests the speaker uses it to indicate that there is relationship between both clauses. In this sense, the collocation has a function similar to English ‘I mean’.

Finally, in all cases of intra-clausal position, the verb appears with an unexpressed subject and without an indirect object. Consequently, it could be possible that *digo* ‘I say’ is emerging as a new construction from the prefab *(yo) digo ALGO* ‘I say something’. Unfortunately, there are not enough tokens in the data to test this hypothesis.

In sum, evidence supporting the existence of the prefab *(yo) digo* ‘I say’ comes from its token frequency, its predominant use to introduce quotes, its tendency to appear in its own intonation unit, and the possibility to occur in different positions within the clause.

5.2. *DIGO YO* ‘I SAY’ (LIT. ‘SAY I’).¹⁰ *Digo yo* represents 10% (5/49) of all 1SG-present clauses, a proportion equal to that of *yo digo* ‘I say’ (5/49: 10%). Certainly, frequency does not provide enough evidence to support the status of *digo yo* as a construction different and independent from the prefab *(yo) digo* ‘I say’; however, it does suggest a tendency for the subject to be expressed after the verb. Moreover, considering VS is not the typical word order for transitive clauses, the subject’s postverbal position relates to a particular discourse function.

Digo yo ‘Say-I’ mostly introduces into the discourse utterances that the speaker has expressed in previous events; nevertheless, complements occur in both preverbal and postverbal position, as shown in 9 and 10, respectively. The only case in which is not clear whether *digo yo* ‘Say I’ has a complement or not is in 11. On the one hand, the clause seems to be truncated by *yo sé que te vas adelantando* ‘and I know you get ahead’; on the other hand, it looks as if the former were indeed an explanation introduced by *digo yo* ‘Say I’.

- (9) Ellos sabían que me –
 .. tenían que respetarme.
 .. y asina he sido siempre,
 le **digo yo**. (311-3B2: 311)
 ‘They know they.. they had to respect me, and I’ve been like that, **I am telling you.**’
- (10) (H) Y por eso **digo yo** que,
 If you do that,
 .. y el mismo patrón de él,
 (H) .. he realizes that. (190-3B2: 177)
 ‘And that’s why **I say** that, if you do that, and his boss himself, he realizes that.’
- (11) [y por eso],
 eh **digo yo** y,
 y yo sé que .. te vas adelantando,
 % % te ayuda mucho,
 en otras palabras. (190-3B2: 105)
 ‘and that’s why, eh **I say**, and I know you get ahead, in other words, it helps you a lot.’

Prosodically, *digo yo* and its complement always appear in different intonation units, as in the examples above. Such a distribution could be related to the fact that speakers always use the collocation to say something about a situation, not to present a speech event per se. That is to say, the collocation does not report actual speech, but it presents an utterance summarizing, or explaining the event being talked about. In 9, for instance, *y asina he sido siempre* is not exactly a quote from the dialogue between the interviewer and the interviewee, but some kind of justification by the latter about his behavior towards his grandchildren.

5.3. COMO (LE) DIGO ‘AS I SAY/TELL (YOU)’. Although this collocation appears only seven times in the data, its morphosyntactic features suggest its prefab status. It represents 14% (7/49) of all 1SG-present tense occurrences of *decir*, 7.5% (7/93) of all 1SG occurrences, and 6% (7/122) of all present tense occurrences. In most of the tokens, the collocation includes an indirect object pronoun referring to whom the linguistic utterance is addressed.

Additionally, *como (LE) digo* ‘As I say/tell (you)’ always appears with unexpressed subjects and it takes clausal complements. However, such complements are not typical because they really do not form a syntactic unit with the verb. Rather, the collocation is intercalated into the information being presented as in 12. Even when *como (LE) digo* ‘As I say/tell (you)’ precedes a whole sentence, it seems to be used as a discourse connector between a sentence and the previous one, as in 13.

- (12) Y yo soy their older sister.
 so yo los --
 yo los cuidaba,
 yo los cuidaba y,
 todavía cuando me casé,
como digo,
 y estaba --
 me enfermaban mis babies y yo,
 y aquí estaba cuidando, (117-1A3)
 ‘and I am their older sister. So I-- I took care of them. I took care of them and,
 even after I got married, as I was saying, and she was -- .. if my babies got sick, I
 was here taking care of them.’
- (13) H: no sabía cuándo parar de jugar.
 F: [Hm]
 H: **[Como] te digo ese hombre**,
 .. ganó.
 Ganaba .. doscientos mil,
 trescientos mil.
 .. Pero se levantaba de la mesa,
 se levantaba pura madre. (156-1A1: 279)
 H: ‘He didn’t know when to stop gambling.’
 F: ‘Hm.’
 H: ‘**I’m telling you**, that man won -- he would win two hundred thousand, three
 hundred thousand, and not even then, he would walk away from the table.’

The relation between *como (LE) digo* ‘As I say/tell (you)’ and the information it refers to seems to be related to the prosodic distribution of both constituents. Except for 13, no elements from the complement share the intonation unit with the collocation. This fact is relevant because the speaker presents the information as if it were already known to the interlocutor, yet it does not appear in the verb’s intonation unit. Thus, prosody suggests the collocation is performing a specific function in the discourse, namely to reactivate a piece of information or to present new information as if it had been previously introduced into the discourse. This process allows the speaker to use a statement as a part of a whole argument or explanation for a particular event. For instance, in 14, the speaker declares she could not speak English fluently as a fact known to her interlocutor, but her intention is to make the point that such circumstances did not impede her and her family from moving forward in life.

- (14) O: Y=,
como digo,
no lo aprendería a hablarlo tan fluently,
.. you know? (117-1A2: 127)
A: Mhm.
O: inglés.
But,
nunca nunca nos atrasamos.
O: ‘And, **as I say**, I might not have learned to speak fluently, you know?’
A: ‘Mhm.’
O: ‘English. But we never got behind in life.’

5.4. *LE(S) DICEN X* ‘THEY CALL IT/THEM X’. One of the clearest prefabs with *decir* ‘to say’ is *le(s) dicen X* ‘they call it/them X’, illustrated in 15. The collocation represents more than half (15/28: 53.5%) of the 3PL_{gen}-present tense occurrences of *decir*, as well as a significant proportion of both 3PL_{gen} subjects (15/58: 26%) and present tense clauses (15/122: 12%).

- (15) [Rastros] **les dicen,**
o **les dicen huellas.**
[no sé]. (214-1B2: 143)
‘Traces, they call them, or they call them footprints. I don’t know.’

As observed in the above example, the collocation has a generic subject, mostly unexpressed (13/15: 86.6%), and an indirect object pronoun referring to an object or place whose name is being asked to the speaker. The name is typically expressed by a bare noun in postverbal position (10/15: 66.6%), as *huellas* ‘footprints’ in 15 above.

On the topic of prosody, the indirect pronoun, the X complement, and the verb share the intonation unit in practically all of the tokens (13/15: 86.6%). Since the complement is the new information introduced by the speaker, such prosodic distribution is only possible if *dicen* ‘they say’ has a low cognitive load. However, the prefab seems to perform a specific discourse function in the discourse; speakers use it to indicate how people from other communities call a particular object, animal, or place. The prefab, then, is a kind of evidential marker, allowing the speaker to present information without being responsible for its accuracy as shown in 16:

- (16) [No sé cómo se] llamarán.
 Te- --
 .. Tepocates les **dicen** en español. (214-1A2: 37)
 ‘I don’t know what their name is. Te-- tepocates, **they call them** in Spanish.’

This prefab is a clear example of how new constructions emerge from previously established ones. In this case, *le(s) dicen X* ‘they call it/them X’ serves as a model for *le(s) decían X* ‘they called it/them X’, as in 17. This prefab appears eight times in the corpus, which represents just 2.6% of the whole data; however, it accounts for 14% (8/58) of all 3PL_{gen} subjects, and 27.5% of all 3PL_{gen}-imperfect clauses (8/29)

- (17) Y hay otras blancas que se dan <X ahí X>.
les dicían tanapes.
 .. unas [blancas]. (10-3A2)
 ‘and there are some white ones that grow there, **they called them tanapes** [turnips]. Some white ones.’

6. CONCLUSIONS. The analysis of the data demonstrates that different prefabs can emerge not only from the same verb, but also from the same collocations. In addition, it shows that not only morphosyntactic features, but also prosodic ones provide a set of operational measures to determine whether a particular collocation constitutes a prefab or, rather, a general structural pattern.

In addition, the analysis provided evidence regarding the role of established prefabs in the development of new ones. Thus, the loss of morphosyntactic variability of *digo* ‘I say’ allows the prefab to move to sentence-final, and even to an internal position, and to develop new semantic and pragmatic nuances. The phenomenon is also observed in the emergence of *le(s) decían X* ‘they called it/them X’, which is related syntactically, semantically, and pragmatically to *le(s) dicen X* ‘they call it/them X’.

REFERENCES

- BARLOW, MICHAEL. 1999. *Monoconc*. Houston, TX: Athelstan.
- BILLS, GARLAND and NEDDY VIGIL. 2008. *The Spanish of New Mexico and Southern Colorado: A linguistic atlas*. Albuquerque, NM: University of New Mexico Press.
- BYBEE, JOAN, 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- BYBEE, JOAN, and CLAY BECKNER. 2009. Usage-based theory. *The Oxford Handbook of Linguistics analysis*, ed. by Bernd Heine and Heiko Narrog, 827–855. Oxford: Oxford University Press.
- BYBEE, JOAN, and RENA TORRES CACOULLOS. 2009. The role of prefabs in grammaticalization: How the particular and the general interact in language change. *Formulaic Language: distribution and historical change* (vol. 1), ed. by Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali, and Kathleen M. Wheatley, 187–217. Amsterdam: John Benjamins.
- CHAFE, WALLY. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: The University of Chicago Press.
- CROFT, WILLIAM. 2001. *Radical construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.
- DAVIES, M. 2002. *Corpus del Español: 100 million words, 1200s–1900s*. Online: <http://www.corpusdelespanol.org>.
- DU BOIS, JOHN; STEPHAN SCHUETZE-COBURN; SUSANNA CUMMING; and DANAE PAOLINO. 1993. Outline of discourse transcription. *Talking data: Transcription and coding in discourse*, ed. by A. Edwards and Martin D. Lampert, 45–89. Hillsdale, NJ: Lawrence Erlbaum.
- ONO, TSUYOSHI, and SANDRA THOMPSON. 1995. What can conversation tell us about syntax? *Alternative linguistics: Descriptive and theoretical modes*, ed. by Philip W. Davis, 213–271. Amsterdam: John Benjamins.
- SCHEIBMAN, JOANNE. 2001. Local patterns of subjectivity in person and verb type in American English conversation. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 61–89. Amsterdam / Philadelphia: John Benjamins.
- SCHEIBMAN, JOANNE. 2002. *Point of view and grammar. Structural patterns of subjectivity in American English conversation*. Amsterdam: John Benjamins.
- TRAVIS, CATHERINE E. 2006. Subjetivización de construcciones: Los verbos ‘cognitivos’ en el español conversacional. *Memoria del VIII Encuentro Internacional de Lingüística en el Noroeste* (vol. 2), ed. by Rosa Ma. Ciscomani Ortiz, 85–109. Hermosillo: UniSon.

NOTES

¹ A period after an IU indicates final transitional continuity; a colon indicates continuing transitional continuity, and a question mark indicates an appeal (cf. Du Bois et al: 1983).

² In Du Bois et al (1983), <X X> indicates unclear speech, and -- indicates a truncated IU. Other features indicated in this paper are truncated word '-', brief pauses '..', medium pauses '...', long pauses '...(N)', speech overlap '[]', and audible exhalation '(H)'.

³ For the examples' translation, punctuation is used in its regular way.

⁴ Generic subjects are indicated with 'gen'.

⁵ It codes whether the clause has a complement and, if so, the type of such constituent. I am using the term 'complement' broadly, to refer to any morphosyntactic structure occurring with the verb to express a whole idea, without focusing on its syntactic function. Thus, coded under this category are not only direct object complements, but also predicative ones, as in examples (3) and (15) respectively.

⁶ Whether the verb and its complement occur in the same intonation unit. Clauses are marked as such when the verb and at least a lexical element of the complement share the same IU.

⁷ Whether the complement is expressed in one or more IU's.

⁸ In fact, token frequency is not an indispensable condition for prefabs, as stated by Bybee (2007:16): 'Prefabs are not especially high in frequency, and yet it is clearly their repetition that has given them their social status of conventionalized and their cognitive status of easily accessed routines.'

⁹ Nevertheless, it is important to mention that, in the discourse, *SBJ-dijo* emerges as a quotative device; it presents speech from previous conversations. More importantly, in such a function, the collocation often appears in its own intonation unit, and it occurs not only before its complement, but also after it, and even shares the complement with another item of *decir*

¹⁰ For this collocation, I will use 'Say I' to distinguish it from both (*yo digo*) and (*yo digo*) 'I say'. In the translations, I will use I SAY.

