

A Database for the Exploration of Spanish Planning

Carlos Gómez Gallo^{1,2}, T. Florian Jaeger^{2,3}, Katrina Furth³

Department of Linguistics, Harvard University¹

Department of Computer Science, University of Rochester²

Department of Brain and Cognitive Science, University of Rochester³

cgallo@fas.harvard.edu

Abstract

We describe a new task-based corpus in the Spanish language. The corpus consists of videos, transcripts, and annotations of the interaction between a naive speaker and a confederate listener. The speaker instructs the listener to MOVE, ROTATE, or PAINT objects on a computer screen. This resource can be used to study how participants produce instructions in a collaborative goal-oriented scenario, in Spanish. The data set is ideally suited for investigating incremental processes of the production and interpretation of language. We demonstrate here how to use this corpus to explore language-specific differences in utterance planning, for English and Spanish speakers.

1. Task-based Multimodal Corpus in Spanish

We present the Spanish Language Fruit Carts corpus based on Aist et al. (2006). This is a video-taped data set of interlocutors instructing a confederate to manipulate objects on a screen. Speakers were free to use any language they chose. The listener uses the mouse to execute the speakers' request and does not give verbal feedback. Hence, the data set is a multimodal corpus formed by interleaving speakers' gestures, spoken instructions, and object manipulations with a mouse.

In each video, a naive speaker and a confederate listener collaborate in executing a common task. The speakers' goal is to replicate a given map by instructing the listener on how to MOVE, ROTATE, or PAINT objects on the computer screen (Figure 1). Since the environment on the computer screen and the reference map differ in the objects' locations, orientations, and colors, the speaker needs to provide elaborate instructions to the listener based on the reference map.

The corpus consists of 120 digital videos of 15 Spanish speakers, undergraduate students, recruited from Universidad de Oriente in Valladolid, Mexico, and Harvard University in Cambridge, MA. Each video ranges from 4 to 8 minutes in duration, with an average of 240 utterances per speaker. The speech was transcribed and annotated by two research assistants.

2. Previous Task-Based Corpora

Corpora can be used to understand how people interpret and produce language-as-action (Clark, 1992). Towards this end, corpora that capture interactive (human-human or human-machine) communication during the execution of a joint activity plays an important role. Various efforts have addressed the need for this type of resource: ATIS (Dahl et al., 1992), TRAINS (Heeman and Allen, 1995), and Map-task (Anderson et al., 1991).

In the ATIS corpus, participants were asked to inquire about air flights reservations, while interacting with a Wizard of Oz (i.e., a human emulating a dialogue system (Kelley, 1985)), or directly with a dialogue system. In the TRAIN

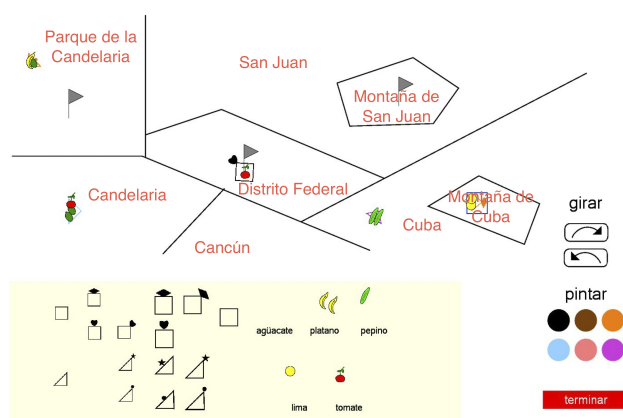


Figure 1: Sample map in privileged view to the naive speaker. Speaker and listener both see the current state of affairs on the computer screen.

corpus, participants were given a task of transporting oranges to factories, making orange juice, and moving orange juice. One of the participants instructed a second one, who played the role of an assistant in carrying out these tasks. In the Maptask corpus, two participants were each given a map which differed slightly from each other: only one of the maps depicted a route, and it had objects in different locations. The participants' task was to successfully draw the route on the map that lacked one. In summary, these corpora provide rich information about task-based collaborative interaction. They are also based on mono-lingual data sets collected in English.

The corpus differs from previous task-based corpora in three ways: a) the type of task that participants execute, b) the data annotation scheme, and c) availability of comparable corpora in multiple languages. In terms of the task, unlike ATIS, our participants learned the goal of their task in a visual, non-linguistic manner. Thus, the task was not testing memory accuracy or capacity; neither was speakers language directly influenced. Unlike Maptask, our task was richer, in that participants had a variety of well defined actions that could be performed (e.g., MOVE, ROTATE, and PAINT). Unlike TRAINS and Maptask, the in-

teraction between participants was limited to having a naive speaker interacting with a confederate listener. Lastly, unlike ATIS and Maptask, our participants shared a common visual ground, and therefore made heavier use of deictic expressions.

In terms of annotation, previous annotation schemes treat each speakers' contribution as an atomic unit (Core and Allen, 1997; Jurafsky et al., 1997; Pineda et al., 2006; Schegloff, 1987). Thus, little is known about how the inner components of each utterance come to be interpreted or produced. Our annotation on the other hand highlights the incremental nature of speakers' utterances. Our corpus also contributes to closing the gap between syntactically annotated corpora (Black et al., 1993; Marcus et al., 1994; Ofizer et al., 2003; Sampson, 2002), and the limited number of semantically annotated corpora (Baker et al., 1998; Ng and Lee, 1996). This makes our data set ideally suited to exploring the incremental production and understanding of speech, during the collaborative execution of a task.

Lastly, there are very few task-based corpora developed in multiple languages. The Fruit Carts corpus is available in two languages: English and Spanish. Hence, our data set provides a comparable corpora that allows the study of these languages in contrast.

2.1. Domain design

The experiment manipulates the complexity of referring expressions while eliciting non-scripted spoken language.

The objects in the experiment consist of a variety of fruits and geometrical figures. Some objects have simple labels (e.g., 'agüacate' 'avocado') and are identical within each set (i.e., all agüacates look the same). Other objects are unlabelled geometrical figures that contrast by: type, size, decoration type, and decoration location (Figure 2 and Table 1). The speaker may produce referring expressions as complex as "El triangulo pequeño con el corazón en la hipotenusa" 'the small triangle with a heart on the hypotenuse', or as simple as "un platano" 'a banana'.

Region names also differ in complexity; for example, "Montaña de Cuba" and "Cuba" (Figure 1). Also there exist points of disambiguation for some regions; for instance, 'Montaña de Cuba' and 'Montaña de San Juan'. In order to avoid ambiguity between these two regions whose labels begin the same way, speakers need to produce the full region name. The regions also have flags which can be used as landmarks, serving as references for MOVE requests.

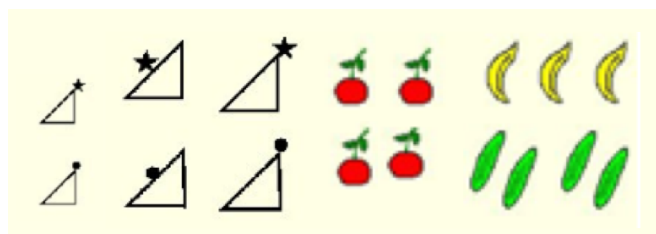


Figure 2: Objects in the experiment.

A typical dialogue, shown in 2, illustrates the variety of referring expressions and non-scripted language in the data.

Object Type	Size	Decoration Type	Decoration Location
Triangle	Small	Star	At the Corner
Square	Big	Diamond	On the Side
Fruit		Dot	none
		Heart	
		none	

Table 1: Corpus 1: Contrasting features of domain objects. Complexity of referring expression was manipulated along four attributes: *Object Type*, *Object Size*, *Decoration Type*, *Decoration Location*.

<p>Speaker> Mueva un triangulo con una estrella en el lado para la Candelaria Listener> (<i>moves triangle to desired goal</i>) Speaker> Allí mismo, luego rotalo uhm Listener> (<i>waits</i>) Speaker> a la izquierda Listener> (<i>starts rotating triangle</i>) Speaker> sigue Listener> (<i>keeps rotating</i>) Speaker> para Listener> (<i>stops</i>)</p>

Table 2: Sample interaction.

Crucially, this experiment manipulates the complexity of referring expressions, while controlling the message that needs to be conveyed. As a result, the corpus is ideally suited to explore questions about how speakers translate a preverbal message into a sequence of orderly linguistic forms. We will now illustrate how this corpus can be used for language production research, that aims at understanding how speakers plan their requests beyond the clausal level.

3. Language-specific properties and inter-clausal planning

Using a similar corpus collected in English (Aist et al., 2006), we investigated how speakers distribute a message across clauses. To illustrate this, consider two, among many, of the choices available to the speaker in realizing a MOVE request. The request can be expressed as a single clause, termed a mono-clausal plan (2), or as two clauses, termed a bi-clausal plan (5). For English, we found in previous work that speakers decided between the two options based on the complexity of the theme expression, but not of the goal expression (Gómez Gallo et al., 2008a).

2. MOVE with an implicit SELECT (Mono-Clausal):

Put [*theme* an apple] [*goal* in Central Park]

5. MOVE with an explicit SELECT (Bi-Clausal):

(a) Take [*theme* an apple]

(b) Move [*theme* it] [*goal* to Central Park]

There is evidence that speakers access both the pronominal adjective and the head noun in parallel, when producing

NPs (Schriefers, 1992). This suggests that there is greater demand of resources when producing a NP that contains a prenominal adjective, than an NP that does not.

In our data, the first NP is the theme expression, as the subject was often omitted due to the imperative forms. Thus, this behavior is consistent with previous studies, which demonstrate that speakers plan the NP (subject) and not the second one (first verb object) at the time of the utterance onset (Lindsley, 1975; Lindsley, 1976). Other studies show that language-specific properties affect planning. Brown-Schmidt and Konopka (2009) find that speakers of English (a prenominal modifying language) plan prenominal modifiers earlier than speakers of Spanish (a postnominal modifying language) plan postnominal modifiers, during NP production. This begs the question of whether modifiers play a role during utterance planning within and across clauses.

4. Sample Study

We used the Spanish language corpus presented in this paper and the English language corpus used in (Gómez Gallo et al., 2008b). We hypothesize that English speakers plan both prenominal modifiers and noun head early, and may experience an extra resource demand from keeping the head and prenominal modifier active in memory. The effects of this interference may be eased by using two clauses instead of one. On the other hand, Spanish speakers may plan head modifiers later, and consequently require resources later. We explore differences in syntactic inter-clausal planning, and argue for the presence of resource/processing limitation when the head and prenominal modifier need to be planned together.

To test this hypothesis, we analyzed 300 and 200 MOVE requests, from 13 English and 15 Spanish speakers, respectively. In both languages, MOVE requests can be realized with a mono-clausal (3) or bi-clausal (4) plan. We coded each utterance for the complexity of both theme and goal expression (word count), whether a prenominal and/or postnominal modifier exists in the theme expression (Figure (3)).

3. Mono-Clausal: Theme modifier

Move [*theme*a small triangle] [*goal*in Central Park]

Mueve [*theme*un triángulo pequeño] [*goal*a la Candelaria]

4. Bi-Clausal:

(a) Take [*theme*a small triangle]

Move [*theme*it] [*goal*to Central Park]

(b) Agarra [*theme*un triángulo pequeño]

Mueve[*theme*lo] [*goal*a la Candelaria]

4.1. Study 1: English

We perform a binary logistic regression model to predict whether speakers choose a bi-clausal over mono-clausal realization based on the following predictors: theme description length, existence of prenominal modifier, and existence of postnominal modifier.



Figure 3: Annotation of two utterances in mono-clausal and bi-clausal realizations.

We hypothesized that English speakers plan heads and head modifiers early, which in turn should affect inter-clausal choice. This is what we find. English speakers decide to distribute a message across two clauses when the theme expression is long, and also when it contains prenominal modifiers (both $ps < .0002$). The effect of postnominal modifiers seems to be weaker and reached only marginal significance ($p < .06$). This suggests that the early planning of the prenominal modifier creates an extra resource demand, which can be dealt with by producing a bi-clausal realization. These effects are independent of each other since they are assessed in the same model and collinearity is controlled for.

4.2. Study 2: Spanish

We perform a second binary logistic regression model, to predict whether speakers choose a bi-clausal over a mono-clausal realization, based on the following predictors: theme description length, and existence of postnominal modifier.

Head modifiers in Spanish are predominately postnominal. Here, speakers may not have to plan modifiers as early as English speakers, and this consequently should not affect early mono- or bi-clausal choices. This is what we find ($p > .18$). Because speakers can plan the head before planning the head modifier, they do not require extra resources, and do not need resort to using a bi-clausal strategy.

5. Summary and Conclusions

The corpus presented is a new resource for the study of language production and comprehension in Spanish. It controls over the conveyed message, while maintaining ecological validity. Here, we have demonstrated that the corpus can be used to explore the manner in which language-specific differences affect utterance planning. Using this corpus, we are able to conclude that head modifier position affects how early speakers plan such modifier, which in turn affects inter-clausal planning.

We plan to include other measures of complexity, beyond expression length. Information content is another measure that estimates complexity. Levy (2006) shows that comprehension difficulty of a word is positively correlated with its information content, or surprisal. Studies in language production have shown that planning is sensitive to the amount of information being processed, and that speakers structure their utterances such that information is distributed relatively uniformly across the utterance, as it unfolds over time (Jaeger, 2006; Jaeger, 2010; Levy and Jaeger, 2007).

This corpus can also be used to explore the relationship between modalities and disfluencies. For instance, previous work has demonstrated that speakers gesture more in contexts when they are producing dispreferred syntactic structures, in English (Cook et al., 2009). Disfluencies have been used to understand how speakers plan their utterances as they correlate with the production of difficult upcoming material (Fox Tree and Clark, 1997). How distant such upcoming material can be remains an open question (Gómez Gallo and Jaeger, 2009).

Spanish speakers recruited for the experiment were bilinguals of either Yukatek Mayan or English. This corpus also allows the comparison between bilingual and monolingual speakers. This is relevant since native grammar in bilinguals is often most affected at the syntax-discourse interface (Polinsky, 2006; Polinsky, 2008; Sorace, 2004).

6. References

- G. Aist, J. Allen, E. Campana, L. Galescu, C. Gómez Gallo, S. Stoness, M. Swift, and M. Tanenhaus. 2006. Software architectures for incremental understanding of human speech. In *Interspeech*.
- A.H. Anderson, M. Bader, E.G. Bard, E. Boyle, G. Doherty, et al. 1991. The HCRC map task corpus. In *LREC*, volume 33, pages 364–370. ELRA.
- C. Baker, C. Fillmore, and J. Lowe. 1998. The Berkeley framenet project. In *ACL*, pages 86–90, NJ, USA. ACL.
- E.W. Black, R. Garside, and G.N. Leech. 1993. *Statistically-driven computer grammars of English: The IBM/Lancaster approach*. Rodopi.
- S. Brown-Schmidt and A. E. Konopka. 2009. Little houses and casas pequeñas: message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*.
- H.H. Clark. 1992. *Arenas of language use*. University of Chicago Press.
- S.W. Cook, T. F. Jaeger, and M. Tanenhaus. 2009. Producing less preferred structures: More gestures, less fluency. In *CogSci*.
- M. Core and J. Allen. 1997. Coding dialogues with the DAMSL annotation scheme. In David Traum, editor, *AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35, Menlo Park, California. AAAI.
- D. Dahl, M. Bates, M. Brown, W. Fisher, K. Hunicke-Smith, D. Pallett, C. Pao, A. Rudnicky, and E. Shriberg. 1992. Expanding the scope of the ATIS task: The ATIS-3 corpus. In *ARPA HLT Workshop*, pages 45–50.
- J. Fox Tree and H. Clark. 1997. Pronouncing “the” as “thee” to signal problems in speaking. *Cognition*, 62:151–167.
- C. Gómez Gallo and F. T. Jaeger. 2009. Early verb choice and fluency as evidence for moderately incremental or possibly limited parallel sentence production. *Talk at CUNY-2009 Conference*.
- C. Gómez Gallo, T.F. Jaeger, and R. Smyth. 2008a. Incremental syntactic planning across clauses. *CogSci*, 40:1294–9.
- C. Gómez Gallo, T.F. Jaeger, and R. Smyth. 2008b. Incremental syntactic planning across clauses. *Proceedings of 30th Annual Meeting of the Cognitive Science Society (CogSci 08)*, 40:1294–9.
- P.A. Heeman and J.F. Allen. 1995. The Trains spoken dialog corpus. CD-ROM. *Linguistics Data Consortium*.
- T. F. Jaeger. 2006. *Redundancy and Syntactic Reduction in Spontaneous Speech*. Ph.D. thesis, Stanford University, Stanford, CA.
- T. F. Jaeger. 2010. Redundancy and reduction: Speakers manage syntactic information density. *In Press*.
- D. Jurafsky, E. Shriberg, and D. Biasca. 1997. Switchboard SWBD-DAMSL Labeling Project Coder’s Manual. Technical report, Draft 13. Technical Report 97-02, Univ. of Colorado Institute of Cognitive Science.
- J. Kelley. 1985. Cal - a natural language program developed with the oz paradigm: Implications for supercomputing systems. In *1st. Conf. on Supercomputing Systems*. ACM.
- R. Levy and T. F. Jaeger. 2007. Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *NIPS*, pages 849–856, Cambridge, MA, December. MIT Press.
- R. Levy. 2006. Expectation-based syntactic comprehension. Master’s thesis, University of Edinburgh, Edinburgh, UK.
- J.R. Lindsley. 1975. Producing Simple Utterances: How Far Ahead Do We Plan?. *CogPsych*, 7(1):1–19.
- J. Lindsley. 1976. Producing simple utterances: Details of the planning process. *JPR*, 5(4):331–354.
- M.P. Marcus, B. Santorini, and M.A. Marcinkiewicz. 1994. Building a large annotated corpus of English: The Penn Treebank. *CL*, 19(2):313–330.
- H.T. Ng and H.B. Lee. 1996. Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach. In *Proc. ACL*, pages 40–47.
- K. Oflazer, B. Say, D.Z. Hakkani-Tur, and G. Tur. 2003. Building a Turkish treebank. *Treebanks: Building and Using Parsed Corpora*, pages 261–277.
- L. Pineda, H. Castellanos, S. Coria, V. Estrada, F. López, I. López, I. Meza, I. Moreno, P. Pérez, and C. Rodríguez. 2006. Balancing transactions in practical dialogues. In *CICLing, LNCS*. Springer Verlag.
- M. Polinsky. 2006. Incomplete acquisition: American Russian. In *Journal of Slavic Linguistics*, pages 191–262.
- M. Polinsky. 2008. Heritage language narratives. In D. Brinton, O. Kagan, and S. Bauckus, editors, *Heritage language education: A new field emerging*, pages 108–156, Hillsdale, NJ. Lawrence Erlbaum.
- G. Sampson. 2002. English for the Computer: The SUSANNE Corpus and analytic scheme. *CL*, 28(1):102–103.
- E.A. Schegloff. 1987. Analyzing single episodes of interaction: An exercise in conversation analysis. *Social Psychology Quarterly*, pages 101–114.
- H. Schriefers. 1992. Lexical access in the production of noun phrases. *Cognition*, 45(1):33–54.
- A. Sorace. 2004. Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition*, 7(02):143–145.