

VARIATION IN INTERROGATIVE ADVERBIALS: *CUÁN*, *QUÉ TAN*, *CÓMO DE*, AND *LO+ADJ./ADV.+QUE*

David Ellingson Eddington
Brigham Young University

ABSTRACT. *Cuán*, *qué tan*, and *cómo de* are used to modify adverbs and adjectives in interrogatives. They are also used in embedded clauses along with *lo+adj./adv.+que*. Instances of these expressions were extracted from the Corpus del Español. In interrogatives, *qué tan* was the most frequent. The idea that *cuán* is archaic or limited to literary usage is not supported by these data. *Cómo de* is extremely infrequent except in Peninsular Spanish. In embedded clauses the frequency of these expressions appear in this order of frequency: *lo+adj./adv.+que* > *qué tan* > *cuán* > *cómo de*. In an experiment, speakers from Spain, Mexico, Puerto Rico, and Venezuela were shown 28 test sentences that contained different adverbial interrogatives. Their task was to choose the expression they preferred. These results correlate highly with the data from the corpus. The choice of adverbial was moderated by gender and age as well.

Keywords. interrogative adverbials; corpus data; experimental approach

RESUMEN. *Cuán*, *qué tan*, and *cómo de* se emplean para modificar adverbios y adjetivos en interrogativas. También se usan en oraciones subordinadas junto a *lo+adj./adv.+que*. En este trabajo se han extraído casos de estas expresiones del Corpus del Español. El más frecuente en interrogativas es *qué tan*. Los datos no apoyan la idea de que *cuán* sea arcaico o esté limitado a un uso literario. *Cómo de* es extremadamente infrecuente excepto en el español peninsular. En cláusulas subordinadas, la frecuencia de estas expresiones aparece en este orden: *lo+adj./adv.+que* > *qué tan* > *cuán* > *cómo de*. En un experimento, hablantes de España, México, Puerto Rico y Venezuela fueron expuestos a 28 oraciones con distintos interrogativos adverbiales. Su tarea consistía en elegir la expresión que preferían. Los resultados muestran una alta correlación con los datos del corpus. Asimismo, la elección de la expresión adverbial estuvo matizada por género y edad.

Palabras clave. adverbios interrogativos; datos de corpus; aproximación experimental

1. Introduction

Exactly how are adjectives and adverbs modified in an interrogative or in an embedded clause in Spanish? In contemporary usage there are three interrogative adverbials that may serve this function: *cuán*, *qué tan*, and *cómo de*.¹

- (1) ¿Cuán frágil le parece su propia existencia?
'How fragile does his own existence seem to him?'
- (2) ¿Qué tan frecuentemente debo cambiar las cuerdas?
'How often should I change the strings?'
- (3) ¿Cómo de contaminada está su mente?
'How polluted is his mind?'

1 Some of these illustrative examples come from the Corpus del Español, or are modified from sentences found therein. Others are provided by the author.

In embedded clauses these three adverbials are joined by *lo+adj./adv.+que*.

- (4) Demuestra lo importante que es la experiencia.
‘It shows how important experience is.’
- (5) Me retó a ver cuán lejos podía yo mear.
‘He challenged me to see how far I could pee.’
- (6) El monitor de distancia te ayudará a determinar qué tan cerca estás.
‘The distance sensor will help you determine how close you are.’
- (7) Así es cómo de peligrosas son las farmacéuticas.
‘That is how dangerous pharmaceutical companies are.’

In their work on quantifiers, Octavio de Toledo y Huerta and Sánchez López (2009) specify *qué tan*, *cuán*, and *cómo de* as the quantifiers that precede adjectives and adverbs in interrogatives and exclamations. As far as their geographical distribution and prestige are concerned, Bello (1847/1859:306) notes the use of both *cuán* and *qué tan* in the Americas in the 19th century, but then, a hundred years later, Bollinger (1946:167) expresses a nagging sense that Latin Americans find *qué tan* as not proper for educated speech. Seco (2000:375), for his part, feels that *qué tan* has an archaic flavor according to his peninsular palate, but he, along with Steel (1999), cite Kany (1951:330) as evidence that it is commonly used in Latin America. La Real Academia Española similarly identifies *qué tan* as an Americanism (Real Academia Española 2005:547). Historically *qué tan* was in general usage from the 14th to the 17th centuries, but fell into disuse in Spain in favor of *cómo de*, while being maintained in Latin America (Sánchez López 2006:48).

So, if *qué tan* is an Americanism that is getting a bad rap in prestigious speech, even in the Americas, how well does *cuán* fare as an alternative? To the ears of some Peninsula speakers (Sánchez López 2006) it sound archaic. The Real Academia Española (2009:1647-8) mark it as an American usage, while Díez Losada (2006), a Spaniard who emigrated to Central America, feels that *cuán* only belongs to the literary register, and is on its way to the grave. If that is the case, what is one to do? He suggests avoiding both of them altogether. One way to achieve this is to express the proposition as a yes/no question rather than as a quantified adjective or adverb. In other words, forget *¿qué tan lejos queda?* or *¿cuán lejos queda?* ‘How far is it?’ and simply ask *¿queda lejos?* ‘Is it far?’ This is one solution that Crispin and Crispin (1972:164) observed Spanish speakers applying.

The other option that Díez Losada suggests is to eliminate the adjective altogether by simply nominalizing it. That is to say, eliminate the *urgente* ‘urgent’ in *cuán urgente* and *qué tan urgente* in favor of its nominal cousin *urgencia*: *¿qué grado (nivel) de urgencia tiene este trabajo?* ‘What degree of urgency does that task have?’ One strategy he does not mention is how matters of size and dimension can be handled. For example, *alto* ‘tall’ and *grande* ‘large’ can often be quantified with the verb *medir* ‘to measure’ or the noun *tamaño* ‘size’ (e.g. *¿Cuánto mide?* ‘What does it measure’ *¿De que tamaño es?* ‘What size is it?’). Of course the other option is *cómo de* which can manifest itself in two ways: *¿Cómo de urgente es?* and *¿Cómo es de urgente?* ‘How urgent is it?’

Nevertheless, Whitley (1986) tackled this sticky issue in the late 20th century. He gave 17 native Spanish speakers (who also spoke English) the task of translating English sentences containing *how+adjective* interrogatives such as “How well do they speak German?” His speakers were from eight different countries. A scant 3 of 391 responses contained *cuán* which lends credence to the notion of its impending demise. Thirteen percent of the translations involved *qué tan* and only 6% some form of *cómo*

de. The remainder included other adjective avoidance maneuvers. For instance the speakers steered away from *frecuente* to express bathing frequency and would instead rephrase the question as *¿Cada cuánto se baña?* ‘How often does he bathe?’ or *¿Se baña a menudo?* ‘Does he bathe often?’ In fact, nominalizing was the strategy used in 30% of the cases.

Whitley also carried out an acceptability study with 13 speakers from six countries. They were given different ways of expressing the same question. He found that, in general, most speakers disliked *cuán* and preferred to nominalize the adjective. As far as preferences are concerned, Latin Americans were more accepting *qué tan*, in contrast to the Spaniards who tended to reject it and were also far more accepting of *cómo de* than were the Latin Americans.

With the exception of Whitley, the issue of interrogative adverbials has not been examined with any amount of detail, and a number of questions are begging to be addressed. For instance, what is their distribution across the Spanish-speaking world? Is *cuán* truly about to be laid to rest? Is *qué tan* actually relegated to Latin American Spanish and *cómo de* limited principally to Peninsular varieties? What about the use of interrogative adverbials in embedded clauses? Does their appearance in that position parallel their interrogative uses? The present study attempts to answer these questions with corpus data as well as with data from an experimental study. It fits in the first and second waves of quantitative studies of linguistic variability (Eckert 2012). That is, the use of the interrogatives is viewed as being linked to different speech communities that in this case are delimited by political boundaries. At the same time it examines the role that individual factors such as gender and age play in the variation.

2. Corpus study

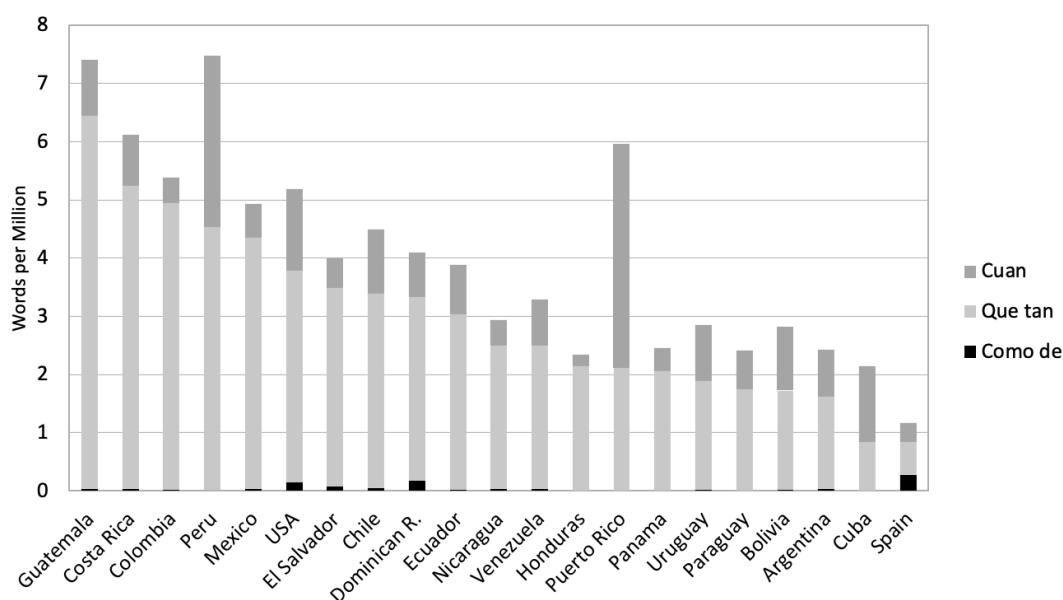
2.1. Corpus study of interrogative adverbial expressions

All data in the corpus study were gleaned from Corpus del Español-Web Dialects (Davies 2016), a corpus of two billion words.² This corpus was created from online sources between 2013 and 2014, 60% of which are blogs. It should be noted before moving on that while corpora are an invaluable source of linguistic data, they are not without their drawbacks (Brown 2015). There are always issues with typographic, compilation, and tagging errors in corpora. A crucial component of the present study is the country the speaker is from, and the possibility exists that the corpus may not have assigned the data from a particular speaker to the correct country. Another factor that cannot be ruled out is that a speaker from one country may be an expatriate who actually resides in another, and for that reason his or her speech would not be representative of the country the person is writing from. The corpus may also contain an inordinate amount of data from the more populated regions of the country which creates problems when it comes to how representative the data may be of the entire country. In a similar way, dividing the data by country is inherently problematic since political boundaries often do not correspond with dialectal boundaries, and a single country typically houses a number of speech varieties. In spite of these issues, corpora provide a wealth of information that can be used to investigate linguistic hypotheses including interrogative adverbials in Spanish.

2 Number of million words in the corpus by country: 169.4 Argentina, 39.3 Bolivia, 66.2 Chile, 166.4 Colombia, 29.5 Costa Rica, 63.2 Cuba, 33.6 Dominican Republic, 52.3 Ecuador, 426.5 Spain, 54.2 Guatemala, 35.1 Honduras, 245.9 Mexico, 32.3 Nicaragua, 22.2 Panama, 107.2 Peru, 32.1 Puerto Rico, 29.7 Paraguay, 36.4 El Salvador, 166.0 USA, 38.7 Uruguay, 98.1 Venezuela

Three searches were carried out on the corpus. The first search included all cases of question sentences beginning with *cuán* or *cuan* followed by an adjective or adverb (e.g. *¿Cuán pronto llegas?* ‘How soon will you get here?’). In the second search, questions beginning with *qué tan* or *que tan* that were followed by an adjective or adverb were obtained (*¿Qué tan atractiva la ves?* ‘How attractive is she to you?’). The last search was somewhat more complex. It included questions starting with *cómo de* or *como de* followed by an adjective or adverb and then either a form of *ser* or *estar* (e.g. *¿Cómo de necesario fue ese castigo?* ‘How necessary was that punishment?’). Another kind of question that belongs to this category consists of questions starting with *cómo* or *como* followed by a form of *ser* or *estar* and then an adjective or adverb (e.g. *¿Cómo somos de libres si no hay derechos humanos?* ‘How free are we if there are no human rights?’). The raw number of tokens obtained in each search was divided by the size of the corpus in each country to standardize the outcome in terms of tokens per million. The graphic results of the searches appear in Figure 1 while the numeric results are found in Appendix 1.

Figure 1. Corpus search results for interrogative adverbials in words per million.



Perhaps the most striking finding is the variation in how often interrogative adverbials of any kind are used in each country. It is highly doubtful that discourse in some countries includes so many more questions with quantified adjectives and adverbs that it does in another country. It is more likely that speakers in countries such as Spain and Cuba, which are on the low end, may prefer alternative ways of constructing interrogatives that do not involve *cuán*, *qué tan*, or *cómo de*. As already discussed, this may entail wording the question in such a way as to eliminate the adjective or adverb altogether. Instead of inquiring *qué tan/cuán rico* ‘how rich’ one may simply ask: *¿Cuánta plata tiene?* ‘How much money does he have?’ In a similar manner the adjective or adverb may be replaced with a corresponding noun. *Lejos* ‘far’ and *frecuente* ‘frequent’ may be expressed with *distancia* ‘distance’ and *frecuencia* ‘frequency’: *¿A qué distancia queda?* ‘What distance is it?’ *¿Con qué frecuencia bebe?* ‘How often does he drink?’ Another option is to verbalize the adjective so that, for example, *importante* ‘important’ becomes *importar* ‘to matter’: *¿Cuán/qué tan importante es?* ‘How important is it’ is expressed as *¿Cuánto importa?* ‘How much

does it matter?’ Unfortunately, there is no straightforward way to quantify how often this approach is taken in the speech of each country by consulting a corpus.

A number of other findings are also apparent. Despite the fact that some Spanish speakers feel that *cuán* has an archaic or literary feel to it, *cuán* appears to be alive and well. It is important to remember that 60% of the data that the Corpus del Español-Web Dialects is comprised of is blogs, which are neither archaic nor generally literary in nature; they actually tend to represent the more informal side of the language register spectrum. What is of particular interest is that interrogatives with *cuán* dominate the other alternatives in Puerto Rico and Cuba, and rival *qué tan* in Peru, Bolivia, Uruguay, and Argentina. This suggests that its literary or archaic feel may not be pan-Hispanic, but restricted to those countries where *cuán* is an infrequently used alternative.

If these three adverbials were in competition, then *qué tan* would surely come out as the undisputed winner. It is the most frequent construction in 19 of 21 countries, including Spain. Additionally, in most countries it dominates the other alternatives by a wide margin. Finally, *cómo de* as an interrogative adverbial registers a mere blip on the chart. The only place it seriously competes with any of the others is in Peninsular Spanish. It is interesting that the Real Academia Española (2009:1648) recognizes *cómo de* as Peninsular, but claims it is also somewhat common in *rioplatense*, and in certain registers in Bolivia, Cuba, and Central America. The present data, in contrast, do not support any significant use in the Americas.

2.2. Corpus study of interrogative adverbials in embedded clauses

While the bulk of the discussion on interrogative adverbials focuses on their use in direct questions, less attention has been paid to their appearance in indirect questions and in embedded clauses such as *No importa cuán lejos esté* ‘It doesn’t matter how far it is.’ In order to study this, relevant data was obtained in a number of corpus searches. Cases of *qué tan*, with and without an accent mark, were found in the context Verb+*qué tan*+Adjective or Adverb (e.g. *Los consumidores pueden determinar qué tan rápido un nuevo producto es adoptado* ‘The consumers can determine how fast a new product is adopted’). Instances of *que tan solo* were eliminated from the counts since they don’t express a degree of loneliness, but instead have the meaning ‘just.’ Tokens of *cuán* and *cuan* that were preceded by a verb and followed by an adjective or adverb were included in the study (e.g. *No importa cuán difícil sean nuestras circunstancias* ‘It doesn’t matter how hard our circumstances may be’). Counts of *cómo de* (and *como de*) were obtained by searching for Verb+*cómo de*+Adjective or Adverb+*Ser* or *Estar* (e.g. *Demuestra cómo de bueno es nuestro espíritu de equipo* ‘It shows how good our team spirit is’), and Verb+*cómo*+*Ser* or *Estar*+*de*+Adjective or Adverb (e.g. *Siendo como eres de clara con todo* ‘Being so clear about everything as you are’).

Although ‘*lo*+Adjective or Adverb+*que*’ cannot serve as an interrogative adverb in direct questions, it does have that function in embedded clauses (Sánchez López 2006; e.g. *Me hirió lo cruel que me trataste* ‘It hurt me how cruelly you treated me’). For that reason it needs to be included in this discussion. Strings comprised of Verb+*lo*+Adjective or Adverb+Verb were counted. However, this included a number of expressions such as *Es lo único que puede funcionar* ‘It’s the only thing that can work’, in which the function of *lo* is to nominalize the adjective or adverb yielding ‘the only thing.’ Since these do not involve quantification of the adjective or adverb, strings containing *mejor*, *peor*, *único*, *mínimo*, *máximo*, *más*, *malo*, *bueno*, *primero*, *último*, and *siguiente* were not included in the counts (‘better, worse, only, least, maximal, most, good, first, last, following’).

The first question to ask is how much the frequencies of these expressions in each country can be correlated between their uses in direct questions and in embedded clauses. As Table 1 indicates, *cuán* and *qué tan* are correlated across their usages, while there is no significant correlation between the uses of *cómo de*, which is not unexpected given its marginal status in most countries.

Many of the same trends observed for direct questions are found here as well (Figure 2. Numeric results appear in Appendix 1). For example, Puerto Ricans have the highest preference for *cuán* while speakers from Spain carve out the largest sliver in the use of the expression *cómo de*. The least frequent use of *lo+adj./adv.+que* belongs to Bolivian speakers which contrasts with Chileans and Spaniards who appear to favor this construction a great deal in embedded clauses.

Question and Embedded clauses	<i>r</i>	<i>p</i>
<i>Cuán</i>	.789	< .001
<i>Qué tan</i>	.433	.050
<i>Cómo de</i>	-.096	.678

Table 1. Correlations between the adverbials in direct questions and in embedded clauses across countries.

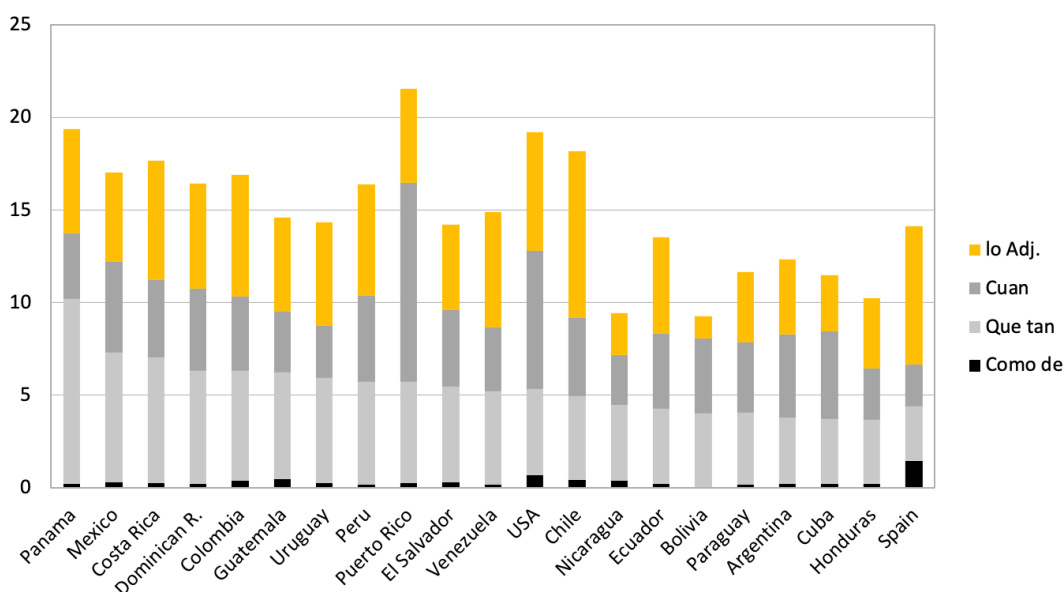


Figure 2. Corpus search results for adverbials in embedded clauses in words per million.

3. An experimental approach to interrogative adverbials

Although the corpus data shed a great deal of light on the distribution of interrogative adverbials across the Spanish-speaking world, it is only one source of data. More definite conclusions can be reached once data from one source have been independently verified with data obtained by other means. For this reason, an experiment was designed to investigate this issue in more depth. One reason for doing this is that the corpus searches are limited to the use of the four adverbial expressions discussed above. However, we know that in addition to the interrogative adverbials included in the corpus search, Spanish speakers express the same concepts in other ways as well, such as

nominalizing the adjective or adverb, or converting the proposition into a yes/no question. In an experiment, the participants were given a free response option so that these alternative methods could be examined. The present experiment was designed to address these matters.

3.1. Participants

Rather than attempt to enroll participants from all Spanish-speaking countries, participants from only four countries were involved in the experiment. Puerto Ricans and Spaniards were chosen since they were much more likely to use *cuán* and *cómo de* (respectively) in the corpus. In like manner, Mexicans are among the highest users of *qué tan*, while Venezuelans represent a country that falls in the middle ground.

A total of 137 participants were included in the study: 71 males, 66 females. They were from the following countries: 26 Mexicans, 44 Puerto Ricans, 39 Spaniards, and 28 Venezuelans. Their mean age was 31.5 with a standard deviation of 10.8. Many more participants began the survey, but had to be eliminated because they were not from the four countries of interest, were under 18 years of age, or replied incorrectly to the lie detector questions (see Test Items).

3.2. Method

Participants were initially recruited on Mechanical Turk from Spain, Puerto Rico, Venezuela, and Mexico, and paid \$1.50 for completing the survey. Since this method did not attract any Puerto Rican participants, their participation was solicited through emails.³ After responding to the biographical questions the participants were randomly assigned to one of two groups. One group saw the A questions and the other the B questions (see Appendix 2). The instructions read as follows:

En el español hay varias maneras de expresar lo mismo. Por ejemplo, para averiguar el nombre de alguien se puede preguntar:

- ¿Cómo te llamás vos?
- ¿Cuál es tu nombre?
- ¿Cómo tú te llamas?
- ¿Cómo te llamas tú?

Seguro que una de las frases la usarías exclusivamente, o por lo menos mucho más que las otras. En esta encuesta verás una serie de frases y tienes que elegir la que dirías si tú estuvieras hablando. ¡Ojo! Es importante que leas las frases completamente y que sigas las instrucciones al pie de la letra. Si no haces esto, la encuesta se va a cortar y no podrás ni volver a empezarla, ni ser pagado por completarla.

‘In Spanish there are various way of expressing the same thing. For example, to find out someone’s name one can ask:

[four ways of asking ‘what is your name’]

I’m sure you’d probably use one of them exclusively, or at least more than the others. In this survey you’ll see a series of sentences and you’ll need to choose which one you’d say if you were speaking. Be careful! It’s important for you to read the sentences completely and to follow the instructions carefully. If you don’t the survey will stop and you won’t be able to start it again, nor be paid for completing it.’

3 Thanks to Sally Delgado, Nick Faraclas, Luís Ortiz, Rosa Guzzardo, Héctor Aponte, Alexandra Morales Reyes, Don Walicek, and Melvin González Rivera for helping find participants in Puerto Rico.

There were 31 questions that were presented in the following format. The participants were asked to choose from among the sentences, or fill in the blank with an alternative way of expressing the same idea if none of the choices suited them:

Estas frases expresan lo mismo, pero ¿cuál de ellas dirías tú?
 ¿Qué tan seguro es comprar por Internet?
 ¿Cuán seguro es comprar por Internet?
 ¿Cómo de seguro es comprar por Internet?
 Ninguna de ellas (Entonces, ¿cómo lo dirías?) _____
 ‘These sentences express the same thing, but which of them would you say yourself?’
 [four ways of asking ‘how safe is it to buy on the internet?’]

Three lie detector questions were also included (see Test Items).

3.3. Test items

The test items (Appendix 2) contained the adjectives *profundo*, *importante*, *seguro*, *difícil*, *fácil*, *probable*, *frecuente*, *diferente*, *feliz*, *fuerte*, *peligroso*, *preparado*, *relevante*, and *serio* (‘deep, important, safe, hard, easy, probable, frequent, different, happy, strong, dangerous, prepared, relevant, serious’). No adverbs were tested. The sentences were taken directly from the Corpus del Español or slightly modified to fit the test format. Each adjective appeared in two direct questions and two embedded clauses. For instance, the following questions are based on *serio* and given with the *qué tan* response:

28 Direct questions

Set A *¿Qué tan serio tratas tu negocio?*
 ‘How seriously do you take your business?’
 Set B *¿Qué tan serio es el riesgo?*
 ‘How serious is the risk?’

28 Embedded clauses

Set A *Eso dependería de qué tan serio es su caso.*
 ‘That would depend on how serious your case is.’
 Set B *Me di cuenta de qué tan serio era el asunto.*
 ‘I realized how serious that matter was.’

Participants were randomly assigned to answer either the A or B set of questions, meaning that they responded to 14 direct questions and 14 embedded clause questions. In this way they were presented each of the 14 adjectives in a direct question sentence and in an embedded clause question. The order of the test questions as well as that of the response choices was randomized. Two different sets of questions were used in order to minimize the influence that any particular sentence could have on the outcome.

In addition to responding to the 28 test questions, three lie detector questions were included. It is an unfortunate reality that many people participating in Mechanical Turk tasks will fill out a survey without actually paying attention to the questions and responses. In order to eliminate responses of this kind, three questions were included all of which had syntactically impossible sentences as choices such as *Eso dependería de tan qué serio tan es su caso* (see Appendix 2). The only valid answer in these cases would be *Ninguna de ellas* ‘None of the above.’ Any other response casts doubt on the

validity of all of that participant's responses, and for that reason they were eliminated from the analysis.

3.4. Results and discussion of interrogative adverbials in direct questions

Figure 3 illustrates just how varied the responses are across the four countries. In order to determine how all the variables affected preferences for adverbials, a series of mixed effects logistic regressions were performed with participants being given random intercepts. The independent variables were the participants' country, age, and gender. One binomial regression compared *qué tan* against the other adverbial expressions combined, another contrasted *cuán* with the other adverbials combined into one value, and a third compared *cómo de* against the other adverbial expressions combined.

Figure 3. Frequency of interrogative adverbial survey responses by country.

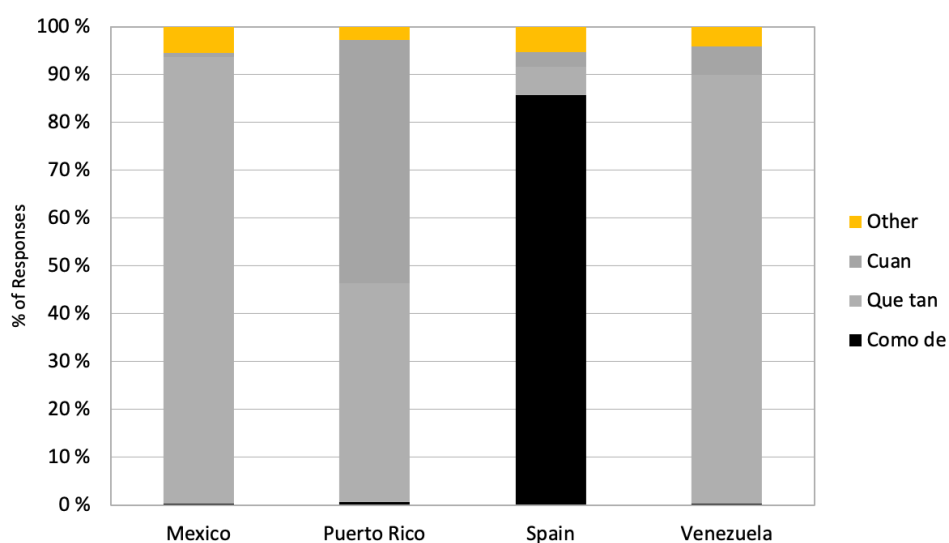


Table 2. *Qué tan* vs. other adverbials combined in questions sentences.

	F	df1	df2	p
AGE	2.432	1	12	.144
SEX	2.432	1	26	.131
COUNTRY	6.104	3	22	.004
SEX * COUNTRY	2.360	3	22	.099
AGE * SEX	3.476	1	12	.086
AGE * COUNTRY	1.418	3	10	.293
AGE * SEX * COUNTRY	2.109	3	10	.160

Table 3. *Cuán* vs other adverbials combined in questions sentences.

	F	df1	df2	p
AGE	3.012	1	23	.096
SEX	.028	1	39	.868
COUNTRY	1.822	3	28	.166
SEX * COUNTRY	4.341	3	28	.012
AGE * COUNTRY	2.389	3	22	.097
AGE * SEX	.065	1	23	.801
AGE * SEX * COUNTRY	4.760	3	22	.011

Table 4. *Cómo de* vs. other adverbials combined in questions sentences.

	F	df1	df2	p
AGE	32.775	1	0	1
SEX				< .0005
COUNTRY	331.095	2	1836	< .0005
SEX * COUNTRY	336.319	2	1836	< .0005
AGE * COUNTRY	1375.944	1	1836	< .0005
AGE * SEX	27.209	1	0	1
AGE * SEX * COUNTRY	1259.731	1	1836	< .0005

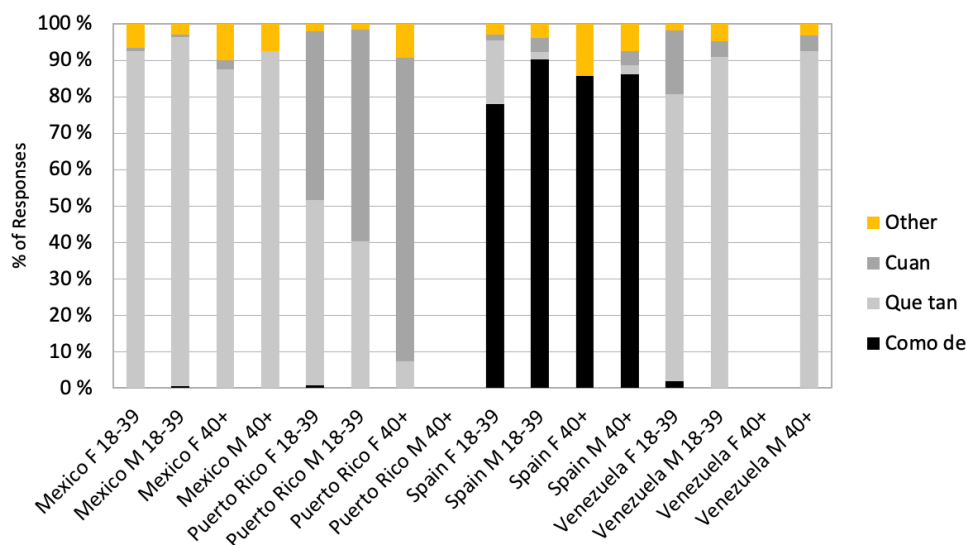
Table 5. Post hoc comparisons of *qué tan* vs. other adverbials combined in questions sentences.

COUNTRY	COUNTRY	Contrast Estimate	SE	test	df	p
Mexico	- Spain	.917	.020	46.644	108	< .0005
Mexico	- Puerto_Rico	.567	.051	11.054	5	< .0005
Mexico	- Venezuela	.056	.029	1.951	7	.094
Spain	- Venezuela	-.861	.031	-27.600	10	< .0005
Puerto_Rico	- Spain	.350	.053	6.657	6	< .001
Puerto_Rico	- Venezuela	-.510	.057	-9.003	5	< .0005

In the *cómo de* and *cuán* analyses the interactions of country, age, and gender were significant, and those data are represented graphically in Figure 4. It would have been ideal to summarize the data in that figure by representing more age groups, but the young mean age of participants (31.5) meant that the bulk of the participants were younger. There were also a number of gaps in age groups without participants in some countries. Both of these would have made a more nuanced graph difficult interpret, and even this binary approach leaves two groups empty (i.e. older Puerto Rican males, and older Venezuelan females). For this reason only two age groups were used.

In any event, the data indicate that Spaniards preferred *cómo de* to a much greater extent than the speakers from the other countries, but to a somewhat smaller degree by the youngest females from that country. In a similar manner, *cuán* is the preferred adverbial by Puerto Ricans, especially by the older females when contrasted with younger Puerto Ricans of both genders, who were more inclined toward *qué tan*.

Figure 4. Frequency of adverbial responses in the experiment by gender, country, and age.



Recently, many voices have been raised against the use of null hypothesis testing to blindly dichotomize variables into those that do and do not influence the results. The idea that the alpha level of 0.05 is special or ideal has been criticized (Amrhein, Greenland & McShane 2019; Amrhein, Trafimow, and Greenland 2019, Cohen 1994). Kline (2004), on his part, actually suggests that an alpha level of 0.1 would be more appropriate in the behavioral sciences. In either case, the fact that some one variable may reach statistical significance while another does not should not be taken to mean that the first one affected the results while the second did not. Instead, a *p* value should be taken as a point at which the data are most compatible, with less compatible, but still influential data falling farther away.

With this in mind, an examination of the results of the *qué tan* analysis reveals that country reaches statistical significance, but there are two other trends. One is an interaction between sex and country ($p = .099$), and between age and sex ($p = .086$). These interactions are evident in Figure 4 which suggests that further investigation into how social factors influence the use of interrogative adverbials in these countries would be a fruitful enterprise, especially when a wider range ages is included.

How do the results from this experiment compare with the frequencies observed in the corpus? There are too few countries to be able to run a correlation by each individual interrogative adverbial. However, the frequency per million of each adverbial in each country from the corpus may be correlated with the percent that each adverbial was chosen by speakers from the four countries. The results indicate that the two studies are highly correlated ($r(10) = .694, p = .012$). This fact bolsters the validity of the findings of both studies.

In addition to being able to chose from among the set responses, the participants could also choose *Ninguna de ellas* (*Entonces, ¿cómo lo dirías* ‘How would you say it then?’). They were not required to fill in the blank, but many did provide alternatives to the set responses. Unfortunately, there were too few to analyze statistically. Nevertheless, of the 82 alternative responses to the question sentences, three were identical to one of the set responses and 29 were reworded sentences. For example, *¿Qué tan probable es el tratamiento para curarme?* ‘How probable is the treatment to cure me?’ was rendered as *¿Cuán efectivo es el tratamiento?* ‘How effective is the

treatment?’ by one participant. Other modifications involved changing the form of address from *usted* to *tú*, replacing *por Tailandia* ‘through Thailand’ with *a Tailandia* ‘to Thailand, and inverting the adjective from *cómo de + adjective es* to *cómo es de + adjective*. Seven open responses converted the test sentence into a question such as *¿___ peligroso es viajar por Tailandia solo?* ‘How dangerous is it to travel through Thailand alone’ into a yes/no question: *¿Es peligroso viajar por Tailandia?* ‘Is it dangerous to travel through Thailand alone?’ This strategy was noted by Crispin and Crispin (1972). The most common change, which occurred in 39 cases, was to eliminate the adverbial expression altogether, by converting the adjective into a noun. *¿___ frecuente pasa eso?* ‘How frequently does that happen?’ is expressed with *frecuencia* instead: *¿Con qué frecuencia pasa eso?* ‘How often does that happen?’ This was a common outcome in Whitley’s study (1986) as well. In four instances the quantified adjective was replaced by an adverb (*frecuente* > *frecuentemente*).

3.5. Results and discussion of interrogative adverbials in embedded clauses

The results in this section deal with adverbials in embedded clauses such as in *Eso dependería de qué tan serio es su caso*. Four different mixed effects regressions were performed with participants given random intercepts. The regressions differed in terms of which adverbial was set as the reference value. In each analysis one adverbial was contrasted with the other three combined. The independent variables were the participants’ country, age, and gender. In the *que tan*, *cómo de*, and *lo+adj./adv.+que* analyses country reached statistical significance, and the effect of country is seen in Figure 5.

Figure 5. Frequency of adverbials in embedded clauses in survey responses by country.

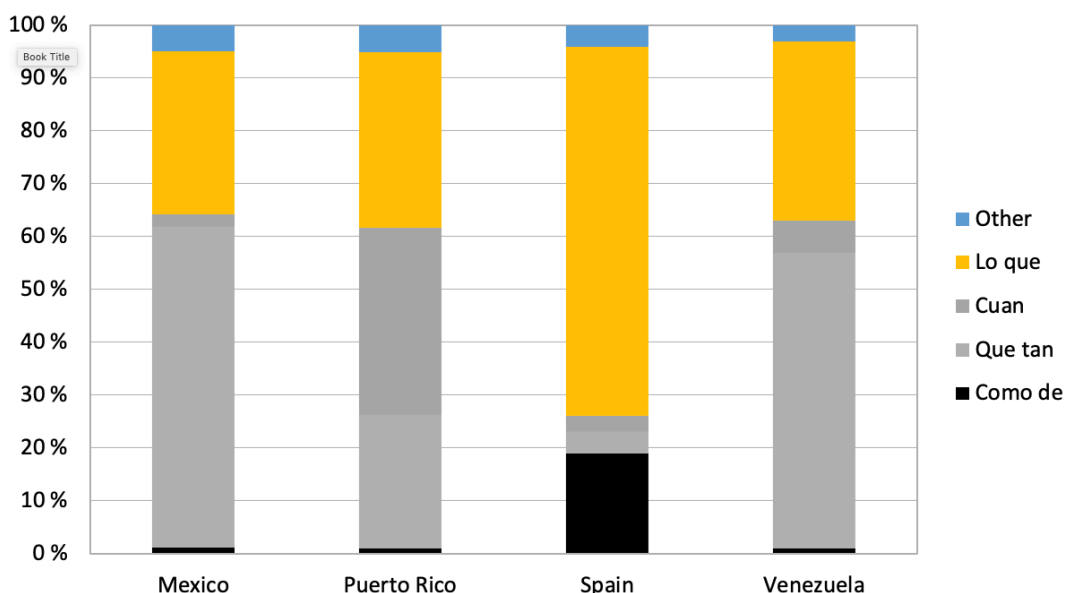


Table 6. *Qué tan vs. other adverbials combined in embedded clauses.*

	F	df1	df2	p
AGE	.156	1	17	.697
SEX	2.508	1	36	.122
COUNTRY	6.250	3	13	.007
SEX * COUNTRY	2.025	3	13	.159
AGE * COUNTRY	4.804	3	6	.054
AGE * SEX	3.138	1	17	.094
AGE * SEX * COUNTRY	2.396	3	6	.174

Table 7. *Post hoc comparisons of qué tan vs. other adverbials combined in embedded clauses.*

COUNTRY	COUNTRY	Difference	SE	test	df	p
Mexico	- Spain	.618	.044	13.966	37	< .0005
Mexico	- Puerto_Rico	.413	.062	6.623	37	< .0005
Mexico	- Venezuela	.105	.057	1.833	6	.120
Spain	- Venezuela	-.512	.042	-12.127	2	.002
Puerto_Rico	- Spain	.204	.049	4.188	51	< .0005
Puerto_Rico	- Venezuela	-.308	.061	-5.0489	7	.001

Table 8. *Cuán vs. other adverbials combined in embedded clauses.*

	F	df1	df2	p
AGE	6432.867	1	0	.959
SEX	14181.395	1	0	.987
COUNTRY	21104.659	3	1836	< .0005
SEX * COUNTRY	21259.937	3	1836	< .0005
AGE * COUNTRY	8749.372	3	1836	< .0005
AGE * SEX	6702.147	1	0	.955
AGE * SEX * COUNTRY	8963.223	3	1836	< .0005

Table 9. *Cómo de vs. other adverbials combined in embedded clauses with age and country as variables.*

	F	df1	df1	p
AGE	.100	1	69	.753
COUNTRY	2.872	3	69	.042
AGE * COUNTRY	5.545	3	69	.002

Table 10. *Cómo de vs. other adverbials combined in embedded clauses with age and sex as variables.*

	F	df1	df1	p
AGE	17.582	1	57	< .0005
SEX	.109	1	57	.742
AGE * SEX	.769	1	57	.384

Table 11. *Cómo de vs. other adverbials combined in embedded clauses with country and sex as variables.*

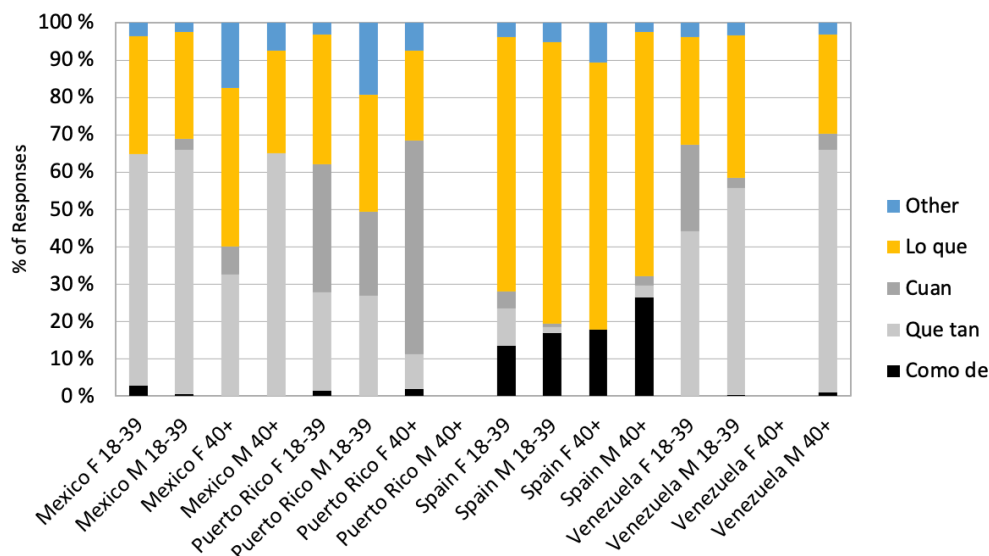
	F	df1	df1	p
SEX	1.721	1	3	.281
COUNTRY	35.306	3	3	.008

Table 12. *lo+adj./adv.+que vs. other adverbials combined in embedded clauses.*

	F	df1	df1	p
AGE	.057	1	21	.813
SEX	.362	1	39	.551
COUNTRY	10.075	3	41	< .0005
SEX * COUNTRY	1.451	3	41	.242
AGE * COUNTRY	3.526	3	23	.031
AGE * SEX	.758	1	21	.394
AGE * SEX * COUNTRY	2.505	3	23	.084

In spite of the large differences between countries, their effect often interacted with other variables (Figure 6). In the *cuán* analysis there was a three-way interaction between age, country, and sex. This is not only evident among the Puerto Ricans, where the oldest females prefer *cuán* to a greater extent than younger speakers of both genders. In Venezuela the youngest females also prefer *cuán* much more than the other groups of Venezuelans.

Figure 6. *Frequency of adverbials in embedded clauses in survey responses by country, sex, and age.*



Given the scarcity of *cómo de* responses in many groups, that analysis resulted in a number of Hessian matrix errors, specifically when all three independent variables were included in the same analysis, and when the interaction of sex and country was included. For this reason, three separate analyses were carried out. Country resulted significant in one, age in another, and age by country in the third. This makes drawing conclusions about the influence of the factors on the use of *cómo de* difficult beyond that fact that it is principally a Peninsular usage.

Country was significant in the analysis of *qué tan*, and it is clear that it is much more prevalent in Venezuela and Mexico than in Puerto Rico and Spain. However, an age by country trend ($p = .054$), as well some influence of age by sex ($p = .094$) can be observed in Figure 6 as well. A similar situation is found for *lo+adj./adv.+que*. Age by country was significant with a trend for age by sex by country also apparent ($p = .084$).

Participants who did not prefer any of the preset responses could provide their own. There were 75 such cases. Fifteen of them were actually duplicates of one the set responses, while 41 were rewrites which involve things such as replacing *quería* with *quisiera*, or removing *yo*. Five additional rewrites inverted the adjective from *cómo de* + adjective *es* to *cómo es de* + adjective. Two participants changed the adjective *frecuente* into *frecuentemente* ‘frequent, frequently.’ In 12 cases the participants avoided the quantifying the adjective altogether, by nominalizing it. For example, *Yo quería saber _____ probable es que yo esté embarazada* ‘I wanted to know how probably it is that I’m pregnant’ was converted into *Yo quería saber cuál es la probabilidad de que yo esté embarazada* ‘I wanted to know what the probability of my being pregnant is.’

A correlation between the adverbials in direct questions and in embedded clauses was tested and found to be significant ($r(14) = .91, p < .0001$). The outlier in this correlation is Spain where *cómo de* was preferred in 86% of the direct question, but only in 19% of the embedded clauses. How do the experimental results for embedded clauses match with those found in the corpus search? They also correlate highly ($r(14) = .739, p = .003$). These correlations help cross validate the findings from the corpus and those resulting from the experiment.

4. Conclusions

There are a number of ways that Spanish speakers have at their disposal to quantify an adjective or adverb. In direct questions they may choose *qué tan*, *cómo de*, or *cuán*, (e.g. *¿Qué tan / cuán / cómo de / importante es* ‘How important is it?’). In embedded clauses these are joined by *lo+adj./adv.+que* (e.g. *No sabía / qué tan / cuán / cómo de / importante era; No sabía lo importante que era* ‘He didn’t know how important it was’). The use of *qué tan* has been assumed to be a characteristic of Latin American varieties, while *cuán* has been brand archaic or literary. A corpus study of these adverbial clauses in direct questions shows that *qué tan* is the most frequently used construction in 19 of 21 countries, Cuba and Puerto Rico being the exceptions where *cuán* dominates. *Qué tan* is followed by *cuán*, and then at a large distance by *cómo de*. The latter is quite infrequent except in Spain. In embedded clauses *lo+adj./adv.+que* is the most common of the four followed by *qué tan*, *cuán*, and then *cómo de*. Their use in direct questions and embedded questions is highly correlated between countries.

While corpora can highlight between country differences, they are not useful for investigating how other factors such as age and gender may play a part in usage differences. For this reason, an experiment was carried out in which speakers from Puerto Rico, Spain, Mexico, and Venezuela were asked to choose which adverbial sounded best to them in a series of direct questions and sentences with the adverbials appearing in embedded clauses. The results of the experiment produced very similar

outcomes to the corpus data for these four countries. As far as *cuán* is concerned it has a vibrant existence in Puerto Rico, but is much more limited in the other four countries. One interesting difference between the corpus and experimental data is that *cuán* was more commonly found in embedded clauses in the corpus than it was chosen in the experiment. That is, speakers may use it in normal speech, but when they are asked to make a conscious choice about how good it sounds to them, they may be influenced by some kind of stigma attached to it. Could the stigma be that it has an archaic feel?

One clear finding is that *cómo de* is not quite, but almost exclusively used and preferred by Spaniards. The Peninsular identity of this adverbial was hinted at in the corpus data, but in the experiment it became obvious that it is the most preferred construction in direct questions, and takes second place behind *lo+adj./adv.+que* in embedded clauses in that country. The fact that the preference for *cómo de* is not as obvious in the corpus data may be due to the influence of Latin Americans living in Spain; country of origin is something that was more tightly controlled in the experiment. As far as *cómo de* is concerned, Latin Americans tend to avoid it in favor of *qué tan* or *lo+adj./adv.+que* (or *cuán* in Puerto Rico). The experiment also demonstrates a number of different ways in which gender and age interact with country. These interactions should be investigated more fully in future research by including participants with a wider range of ages, studying the use of these adverbials in other countries, and including social variables such as social class, education, and rural vs. urban as possible influencing factors.

In the present experiment, participants choose from among preset sentences. Some chose to provide their own response rather than one of the preset ones, and in a number of cases they converted the open question in to a yes/no question or avoided using the adjective in their response by substituting it with a related verb or noun instead. In an experimental paradigm where open responses are solicited rather than preset responses, these kinds of adjective avoidance strategies may become more frequent.

David Ellingson Eddington
Brigham Young University
4064 JFSB
Provo, Utah 85602, USA
eddington@byu.edu

References

- Amrhein, Valentin, Sander Greenland, and Black McShane. 2019. Retire statistical significance. *Nature* 567. 305-307. <https://doi.org/10.1038/d41586-019-00857-9>
- Amrhein, Valentin, David Trafimow, and Sander Greenland. 2019. Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician* 73. 262-270. <https://doi.org/10.1080/00031305.2018.1543137>
- Bello, Andres. 1847/1859. *Gramática de la lengua castellana destinada al uso de los americanos*. Caracas: V. Espinal.
- Bollinger, Dwight. 1946. Que tan, qué tanto. *Hispanic Review* 14.167-169. <https://doi.org/10.2307/470051>
- Brown, Earl K. 2015. On the utility of combining production data and perceptual data to investigate regional linguistic variation: The case of Spanish experiential

- gustar “to like, to please” on Twitter and in an online survey. *Journal of Linguistic Geography* 3(2). 47–59. <https://doi.org/10.1017/jlg.2016.1>
- Cohen, Jacob. 1994. The earth is round ($p < .05$). *American Psychologist* 49, 997-1003. <https://doi.org/10.1037//0003-066X.49.12.997>
- Crispin, Ruth K. and John Crispin. 1972. *Progress in Spanish: Grammar an composition in the second year*. Chicago: University of Chicago Press.
- Davies, Mark. 2016. *Corpus del español: Two billion words, 21 countries*. <http://www.corpusdelespanol.org/web-dial/>
- Díez Losada, Fernando. 2006. La tribuna del idioma: ¿Qué tan...? y ¿cuán...? en español. *La Nación*, Dec. 10. <https://www.nacion.com/ciencia/la-tribuna-del-idioma-que-tan-y-cuan-en-espanol/AKC4A6WVDJHC5O4PP4TBJYJ4R4/story/>
- Eckert, Penelope (2012). Three waves of variation study: The emergence of meaning in the study of variation. *Annual Review of Anthropology*, 41,87-100. <https://doi.org/10.1146/annurev-anthro-092611-145828>
- Jamovi project (2018). *Jamovi (Version 0.9)* www.jamovi.org
- Kany, Charles E. 1951. *American-Spanish syntax*, 2nd ed. Chicago: University of Chicago Press.
- Kline, R. B. 2004. *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington D.C.: American Psychological Association. <https://doi.org/10.1037/10693-000>
- Octavio de Toledo y Huerta, Álvaro S. and Cristina Sánchez López. 2009. Cuantificadores II. Los cuantificadores interrogativos y exclamativos. In Concepción Company Company (dir.), *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal*. México: Universidad Nacional Autónoma de México – Fondo de Cultura Económica, vol. 2, pp. 963-1072
- Real Academia Española. 2005. *Diccionario panhispánico de dudas*. Madrid: Santillana Ediciones Generales.
- Real Academia Española and Asociación de Academias de la Lengua Española. 2009. *Nueva gramática de la lengua española*, vol. 1. Madrid: Espasa Libros.
- Sánchez López, Cristina. 2006. *El grado de adjetivos y adverbios*. Madrid: Arco Libros.
- Seco, Manuel. 2000. *Diccionario de dudas y dificultades de la lengua española*, 10th ed. Madrid: Espasa Calpe.
- Steel, Brian. 1999. *Breve diccionario ejemplificado de americanismos*. Madrid: Arco/ Libros.
- Whitley, M. Stanley. 1986. ‘How’: The missing interrogative in Spanish. *Hispania* 69.82-96. <https://doi.org/10.2307/341142>

Appendix 1

Frequency per million of interrogative adverbials in question sentences.

	Cuán	Qué tan	Cómo de
Argentina	0.8	1.59	0.035
Bolivia	1.09	1.7	0.025
Chile	1.1	3.34	0.045
Colombia	0.44	4.92	0.018
Costa Rica	0.88	5.2	0.034
Cuba	1.3	0.84	0
Dominican R	0.77	3.15	0.178
Ecuador	0.84	3.02	0.019
El Salvador	0.52	3.4	0.082
Guatemala	0.96	6.41	0.037
Honduras	0.2	2.14	0
Mexico	0.58	4.31	0.041
Nicaragua	0.43	2.47	0.031
Panama	0.4	2.06	0
Paraguay	0.67	1.74	0
Peru	2.94	4.53	0.009
Puerto Rico	3.85	2.11	0
Spain	0.32	0.57	0.272
Uruguay	0.96	1.86	0.026
USA	1.4	3.64	0.145
Venezuela	0.8	2.46	0.031

Frequency per million of interrogative adverbials and *lo+adj./adv.+que* in embedded clauses.

	Qué tan	Cómo de	Lo Adj.	Cuán
Argentina	3.61	0.13	4.85	4.45
Bolivia	3.96	0.05	1.88	4.04
Chile	4.55	0.26	10.06	4.23
Colombia	5.93	0.35	7.37	3.98
Costa Rica	6.79	0.1	7.91	4.16
Cuba	3.5	0.17	4.73	4.73
Dominican R.	6.11	0.21	6.77	4.42
Ecuador	4.05	0.1	6.34	4.08
El Salvador	5.18	0.19	5.81	4.11
Guatemala	5.75	0.39	6.45	3.28
Honduras	3.45	0.06	5.21	2.79
Mexico	6.98	0.2	5.79	4.91
Nicaragua	4.07	0.31	3.8	2.69
Panama	9.96	0.04	6.77	3.54
Paraguay	3.89	0.07	4.8	3.79
Peru	5.54	0.14	6.94	4.65
Puerto Rico	5.47	0.19	6.71	10.75
Spain	2.95	0.54	8.83	2.29
Uruguay	5.66	0.21	6.59	2.84
USA	4.64	0.35	7.66	7.45
Venezuela	5.03	0.12	7.21	3.45

Appendix 2

Test questions

Example question set for direct questions:

Estas frases expresan lo mismo, pero ¿cuál de ellas dirías tú?

¿Qué tan seguro es comprar por Internet?

¿Cuán seguro es comprar por Internet?

¿Cómo de seguro es comprar por Internet?

Ninguna de ellas (Entonces, ¿cómo lo dirías?) _____

Example question set for indirect questions:

Estas frases expresan lo mismo, pero ¿cuál de ellas dirías tú?

Eso dependería de qué tan serio es su caso.

Eso dependería de cuán serio es su caso.

Eso dependería de cómo es de serio su caso.

Eso dependería de lo serio que es su caso.

Ninguna de ellas (Entonces, ¿cómo lo dirías?) _____

Direct questions with *qué tan* version

A ¿Qué tan profundo llega el agujero de conejo?

B ¿Qué tan profundo puede bucear?

A ¿Qué tan importante es la alimentación para practicar algún deporte extremo?

B ¿Qué tan importante consideras la tecnología para la educación secundaria?

A ¿Qué tan seguro es comprar por Internet?

B ¿Qué tan seguro se siente en su ciudad o municipio?

A ¿Qué tan difícil puede ser?

B ¿Qué tan difícil es para usted mantenerse al día en nuevas tecnologías?

A ¿Qué tan fácil sería implementar un acuerdo de ese tipo?

B ¿Qué tan fácil es conseguir drogas en Europa?

A ¿Qué tan probable es que los filtros solares no den una reacción alérgica?

B ¿Qué tan probable es el tratamiento para curarme?

A ¿Qué tan frecuente pasa eso?

B ¿Qué tan frecuente le aparece la irritación debajo de su anillo?

A ¿Qué tan diferente puede ser la nueva versión?

B ¿Qué tan diferente es tu vida ahora, en una gran ciudad?

A ¿Qué tan feliz crees que puedes ser?

B ¿Qué tan feliz ha sido usted durante su vida?

A ¿Qué tan fuerte es nuestra relación?

B ¿Qué tan fuerte estás para enfrentar la tentación?

A ¿Qué tan peligroso es viajar por Tailandia sólo?

B ¿Qué tan peligroso sería esto en la sociedad en la que vivimos?

A ¿Qué tan preparado te sentiste para estudiar en otro país?

B ¿Qué tan preparado está el país para enfrentar esta enfermedad?

A ¿Qué tan relevante es para nuestras vidas?

B ¿Qué tan relevante es reflexionar sobre lo que fue su historia?

A ¿Qué tan serio tratas tu negocio?

B ¿Qué tan serio es el riesgo?

Embedded clauses with *qué tan* version

- A Te mostraré qué tan profundo es el Agujero del Conejo.
 B Otros factores afectan qué tan profundo viajan las partículas.
 A Este artículo demuestra qué tan importante es salvar y divulgar este tipo de hallazgos
 B Uno debería preguntarse qué tan importante es la esperanza.
 A Quisiera saber qué tan seguro es el envío de cheques.
 B Averigua qué tan seguro es tu navegador web.
 A No importa qué tan difícil sea la situación
 B No pienses en qué tan difícil es alcanzar la meta.
 A Me pregunto qué tan fácil es conseguir un título.
 B Verá qué tan fácil es cautivar y encantar a una mujer.
 A Yo quería saber qué tan probable es que yo este embarazada.
 B Quisiera saber qué tan probable es que falle la operación.
 A Es importante que lleven un registro para saber qué tan frecuente ocurre.
 B Me gustaría saber qué tan frecuente es el robo en ese barrio.
 A Nos daremos cuenta de qué tan diferente es la naturaleza de la verdadera realidad.
 B Explícale qué tan diferente es hacer las cosas en Internet ahora.
 A Imagínate qué tan feliz estoy despues del divorcio.
 B La calidad de vida es la que determina qué tan feliz es una población
 A Quiero despertar y ver qué tan fuerte soy y hasta donde puedo llegar.
 B Él nos pone a prueba para ver qué tan fuerte es nuestra fe.
 A Por favor explíqueme qué tan peligroso es el hematoma.
 B Desconozco qué tan peligroso resultaría el tratamiento.
 A Todo depende de qué tan preparado estás.
 B No importa de donde venga o qué tan preparado venga.
 A No sé qué tan relevante es esto.
 B Hay que analizar qué tan relevante resultó el cambio.
 A Eso dependería de qué tan serio es su caso.⁴
 B Me di cuenta de qué tan serio era el asunto.

Lie detector questions

Estas frases expresan lo mismo, pero ¿cuál de ellas dirías tú?

- No sé qué tan cuan relevante es esto.
 No sé que cuan relevante tan es esto.
 No sé como es de qué tan relevante esto.
 No sé lo que relevante es esto.

Ninguna de ellas (Entonces, ¿cómo lo dirías?) _____

Estas frases expresan lo mismo, pero ¿cuál de ellas dirías tú?

- Eso dependería de tan que serio tan es su caso.
 Eso dependería de cuan que serio es su caso.
 Eso dependería de que como es tan serio su caso.
 Eso dependería de que lo serio que es que su caso.

Ninguna de ellas (Entonces, ¿cómo lo dirías?) _____

Estas frases expresan lo mismo, pero ¿cuál de ellas dirías tú?

- Por favor explíqueme qué tan peligroso como es el hematoma.
 Por favor explíqueme cuan tan peligroso es el hematoma.

4 This test item was incorrectly coded in the online survey so the results of the choices made by participants could not be included in the analysis.

Por favor explíqueme como de que peligroso es el hematoma.

Por favor explíqueme lo peligroso es como el hematoma.

Ninguna de ellas (Entonces, ¿cómo lo dirías?) _____

Access to full experimental data related to this article can be found here:

<http://linguistics.byu.edu/faculty/deddingt/Embedded%20interrogatives.csv>

<http://linguistics.byu.edu/faculty/deddingt/Interrogatives.csv>