Behav Res
DOI 10.3758/s13428-014-0513-8

Normative data for the 56 categories of Battig and Montague (1969) in Spanish

Alejandra Marful · Emiliano Díez · Angel Fernandez

Abstract Knowledge of specific characteristics of verbal material is imperative in cognitive research, and this need calls for periodical updating of normative data. With this aim, and considering that the most recent Spanish-language category norms for adults date back to more than 30 years ago, and that they do not include some very common categories, a new normative study was conducted. In this study, production data for exemplars in the 56 categories of Battig and Montague (Journal of Experimental Psychology, 80, 1-46, 1969) were collected from a pool of 284 young adults who were native speakers of Spanish using an exemplar production task. With the goal of providing a useful tool for cognitive research to be conducted with Spanish-speaking samples, indices of frequency, rank, and lexical availability for the exemplars of each category are provided in a computerized database. The norms described are available for downloading as supplemental material with this article.

Keywords Spanish category norms · Category exemplars · College students

Introduction

Category norms are highly valuable instruments in cognitive research. Normative studies of categorical representation can provide significant information regarding the content and the organization of stable knowledge structures in our semantic memory (e.g., Barsalou, 1985). They can also be of help in capturing age-related (e.g., Carneiro, Albuquerque, & Fernandez, 2008; Howard, 1980; Price & Connolly, 2006; Yoon et al., 2004) or cross-cultural (e.g., Brown, & Davies, 1976; Yoon et al., 2004) variations in semantic representations. Furthermore, categorical norms provide measurable indices of category exemplars that can be very useful in several areas of experimental research. The utilization of this type of quantitative information in memory investigation, for instance, has been common in studies related to false memories (e.g., Carneiro, Albuquerque, & Fernandez, 2009; Smith, Ward, Tindell, Sifonis, & Wilkenfeld, 2000), retrieval-induced forgetting (e.g., Anderson, Bjork, & Bjork, 1994; Saunders, Fernandes, & Kosnes, 2009), and directed forgetting (e.g., Abel & Bauml, 2013; Zacks, Radvansky, & Hasher, 1996), among others.

The most widely used category norms were created by Battig and Montague (1969). In their study, a large sample of university students (N=442) were provided with the names of 56 categories and they were asked to write as many category members for each item as possible, over 30 seconds. The responses were then processed to compute the total frequency of unique exemplars in each category and their average output position. Published many years ago, they continue to be the reference of choice for many researchers working with English-speaking participants, and similarly constructed norms are also used in other languages. A consensus about the use of particular normative databases and continuous use of the same materials over decades has definitely strengthened research in this area. The most obvious benefit has been a standardization of procedures and a facilitation of direct comparisons among the results obtained by different researchers. Nonetheless, certain properties of experimental stimuli, especially in the verbal domain, may change over time and generations due to variations in use and availability

Electronic supplementary material The online version of this article (doi:10.3758/s13428-014-0513-8) contains supplementary material, which is available to authorized users.

A. Marful (✉)
Departamento de Psicología, Universidad de Jaén, Jaén, Spain
e-mail: mmarful@ujaen.es

E. Díez · A. Fernandez
INICO, Universidad de Salamanca, Salamanca, Spain

Published online: 27 August 2014

among the population at large. Aware of this possibility, while working on associative relationships among words, Jenkins and Palermo (1965) recommended carrying out normative studies from time to time. In fact, Battig and Montague's (1969) efforts in presenting their influential norms for categories and their exemplars were driven, in part, by B.H. Cohen's suggestion to update rather than reprint his previously published norms (Cohen, Bousfield, & Whitmarsh 1957). In the same spirit, and with the aim of including data for additional categories, Van Overschelde, Rawson, and Dunlosky published an updated and expanded version of Battig and Montague's norms in 2004.

The passage of time is not the only factor to influence the validity of categorical normative data. In addition, and in spite of a certain degree of universality in our representations of conceptual and world knowledge, these kinds of norms may be constrained by the peculiarities of specific geographic, linguistic, and cultural contexts. Consequently, researchers from different countries have opted for developing very specific category norms, directly obtained for specific languages, dialects, or populations of speakers. For example, in the case of the English language, norms have been developed not only in the USA (Van Overschelde et al., 2004), but also in Australia (Casey & Heath, 1988) and New Zealand (Marshall & Parr, 1996). In addition, similar materials have been developed for adult speakers of Dutch (Ruts et al., 2004), French (Bueno & Megherbi, 2009), Italian (Boccardi & Cappa, 1997), and Portuguese (Pinto, 1992), to name a few European-origin languages.

In Spanish, there are two relevant studies inspired by the seminal work of Battig and Montague (1969), applying their general procedure of data collection, and using many of their category cues. The first norms for adults were developed by Pascual and Musitu (1980), who collected the responses of 125 Spanish university students to a set of 40 categories (36 from Battig & Montague and four newly created). Approximately at the same time, Soto, Sebastián, García, and del Amo (1982) conducted their own normative study with a much larger sample of Spanish university students, and included some additional categories, resulting in a total of 45 (32 from Battig & Montague and 13 that were new). Because of their high demand, they were reprinted years later, under a slightly different title (Soto, Sebastián, García, & del Amo, 1994), and they continue to be the most frequently used norms of their kind by cognitive and educational researchers in Spain. More recently, category norms for children between the ages of 6 and 13 years were developed by Goikoetxea (2000).

Therefore, considering that the most recent normative materials for adults date from more than 30 years ago, cognitive research on psychology and allied fields would benefit from updated category-norms for the Spanish language. As discussed above, updating after a period of years seems very appropriate for verbal norms mainly for two reasons: first, there could be changes in the frequency and order position of particular exemplars, and, second, new elements could be assigned to the categories due to changes in everyday contexts (e.g., new exemplars in the category *a kind of money*, as Spain has had a new monetary system since 2001).

An additional motivation for the present study was the fact that the existing norms for Spanish adults (Pascual & Musitu, 1980; Soto et al., 1982), although clearly influenced by the study of Battig and Montague (1969), both in procedural aspects and contents, do not include the same 56 categories. Thus, the set of categories tested by Pascual and Musitu have a 64 % overlap with the ones in Battig and Montague (1969), while in the case of Soto et al. (1982) the overlap is 57 %. In absolute terms, this lack of coincidence in the materials may be unimportant, but because the work of Battig and Montague has played such a seminal role in shaping research related to categorical processing in different ways, availability of the whole set of the original categories might facilitate the work of cognitive researchers. For example, picture processing using Snodgrass and Vanderweart's (1980) stimuli has been of interest to a number of researchers working with Spanish-speaking participants (e.g., Cuetos, Ellis, & Alvarez 1999; Moreno-Martínez & Montoro, 2012; Sanfeliu & Fernandez, 1996). Because these drawings represent exemplars originally normed by Battig and Montague, similar category norms would allow for richer information about the items and, therefore, for finer control and manipulation of stimuli. It should also be noted that cross-linguistic and bilingualism studies involving the Spanish language are common (e.g., Macizo, Bajo, & Paolieri, 2012), and the use of comparable categorical norms would contribute to the desired homogeneity in stimuli across languages.

In summary, the need for periodical revision, linguistic specificity, and homogeneity of materials, together with the relevance of categorical norms in current cognitive research, justify the development of updated and expanded adult category norms for the Spanish language. Here we describe methodological aspects and present the resulting normative material.

Method

Participants

A sample of 284 psychology undergraduates from the University of Salamanca, Spain, participated voluntarily in this study to fulfill a course requirement. The majority of them (89 %) were female, and all were native Spanish speakers from different geographical regions in Spain. Twenty-four participants reported being bilingual, with Spanish as their first language.

Materials

The 56 category labels that were used as cues in the exemplar-production task were closely adapted to Spanish from the ones used by Battig and Montague (1969). In order to have simpler, more uniform, and linguistically appropriate labels, and following the labeling tradition in earlier normative studies in Spanish, plural forms of the original expressions were used. Other than that, the resulting labels were straight translations of the original ones, with the exception of the category *States*, which was translated as *Provinces* due to the different territorial organization in Spain. A list of the category labels used is shown in Table 1, together with the original labels and category numbers given by Battig and Montague, to facilitate consultations and comparisons between the Spanish and the English norms.

Procedure

The general procedure of the present study was very similar to the one described in Battig and Montague (1969), with the exception that computers were used for both the presentation of materials and the collection of responses (see Van Overschelde et al., 2004, for a similar data-collection approach). A custom-written software program allowed for the presentation of category cues and the recording of typed responses.

Data-collection sessions were conducted in groups of 17 20 persons in a large computer room, with participants performing the task individually. Each experimental session lasted approximately 1 hour and 15 minutes. At the beginning of each session, participants were informed about the nature of the task and received instructions on how to perform it. They were told that the labels of several categories would be presented on the screen of their computers and that their task was to write, for each category, as many of its exemplars as they could, in the order they came to mind. Participants received general instructions on how to type their responses using the computer keyboard, and practiced generating and typing exemplars for a category (computer components) not included in the final norms.

After the practice trial, a sequence of 56 category labels was presented, with each cue appearing on the top of the computer screen together with blank slots where participants could write their responses during the following 60 seconds. A message on the top of the screen was presented 45 seconds into each trial, informing participants that the remaining time for typing exemplars for the current category was 15 seconds. Category labels were presented in random order for each participant, and there was a 10-second interval between the end of one exemplar-generation trial and the beginning of the next one.

Results

Of the 284 participants, only eight did not respond to one of the categories. The rest of the participants produced at least one exemplar from each cued category.

The responses originally typed by the participants were submitted to an editing process by the first author. First of all, typographical errors and spelling mistakes, potentially affecting 6 % of the responses, were corrected when unambiguous words could be clearly identified. As a result, only 1 % of the original entries were excluded because of being illegible or ambiguous (i.e., they could be interpreted as two or more different lexical entities). Next, and following the standard procedure in normative studies of this kind, variations of the same response were grouped into a unique form: numerals and abbreviations were converted to their spelt out equivalents (e.g., *2* converted to *dos*); variations of trademarks were converted to their most frequent format (e.g., *cocacola* converted to *COCA-COLA*); and accepted variations in the written form of the same exemplar were grouped under the most frequent form (e.g., *kiwi* and *kivi* both coded as *kiwi*). Finally, exemplars repeated by a participant within the same category were eliminated. Other than that, the literal responses of the participants were taken into account, even when an exemplar was manifestly an inappropriate response for the cued category (e.g., *bird* as a four-footed animal).

In constructing the final database, indices of total frequency, rank, lexical availability, and partial frequency were calculated for each exemplar in each category. Total frequency represents the proportion of the 284 participants who produced a given exemplar in a specific category. Rank of the exemplar is an estimation of its average output position in the production task, with lower scores indicating earlier output positions.

Lexical availability is a numerical index that measures how easily an exemplar is generated as a member of a category (Hernández-Muñoz, Izura, & Ellis, 2006; Izura, Hernández-Muñoz, & Ellis, 2005). This index was calculated for each word from the equation developed by López Chávez and Strassburguer-Frías (e.g., López-Chávez & Strassburguer-Frías, 2000) on the basis of the position that the word occupies in the list, the number of participants who generate the word at these positions, and the lowest position where the word was produced. Thus, this index allows the combination of both frequency and rank values, with higher values of lexical availability indicating that the exemplar was produced earlier and more frequently:

$$D(P_j) = \sum_{i=1}^{n} e^{\left[-2{,}3\times\left(\frac{i-1}{n-1}\right)\right]} \times \frac{f_{ji}}{I_1}$$

$D(P_j)$ is the lexical availability of the word j for a given category; I_1 is the total number of participants, i is the position of word j in the list; f_{ji} refers to the number of participants who

Table 1 Categories, Mean Number of Responses, Dispersion, Correlation, Reliability Values, and Hellinger Affinity Values in the 32 Categories that were Common to Soto et al. (1982) and the present study

#	Categories	Mean Responses	Dispersion	Correlation		Reliability	Hellinger Affinity
1	Piedras preciosas (A precious stone)	5.32	20.41	-.32	**	1	.90
2	Unidades de tiempo (A unit of time)	10.40	21.73	-.33	***	.989	.95
3	Parientes (A relative)	13.42	28.88	-.34	***	.993	.96
4	Unidades de longitud (A unit of distance)	7.83	18.23	-.24	**	.996	.94
5	Metales (A metal)	6.28	20.99	-.36	***	.998	.95
6	Material de lectura (A type of reading material)	7.37	6.40	-.23	***	.992	.83
7	Graduaciones militares (A military title)	6.15	16.04	-.34	***	.997	
8	Animales de cuatro patas (A four-footed animal)	12.01	23.05	-.35	***	.995	
9	Tejidos (A kind of cloth)	7.86	14.13	-.30	***	.996	.79
10	Colores (A color)	13.32	31.79	-.54	***	.996	
11	Utensilios de cocina (A kitchen utensil)	10.67	11.97	-.29	***	.995	.89
12	Edificios religiosos (A building for religious service)	5.82	13.44	-.23	*	.996	
13	Elementos gramaticales (A part of speech)	8.49	9.42	-.30	***	.974	
14	Muebles (An article of furniture)	9.72	14.84	-.31	***	.998	.88
15	Partes del cuerpo humano (A part of the human body)	15.19	20.74	-.37	***	.988	.93
16	Frutas (A fruit)	12.19	27.93	-.38	***	.996	.92
17	Armas (A weapon)	8.29	12.32	-.23	**	.995	.89
18	Cargos políticos (An elective office)	6.35	8.71	-.18	**	.996	
19	Tipos de viviendas (A type of human dwelling)	8.06	11.50	-.22	**	.991	
20	Bebidas alcohólicas (An alcoholic beverage)	8.69	17.03	-.28	***	.998	
21	Países (A country)	13.45	20.54	-.34	***	.994	
22	Delitos (A crime)	7.45	6.29	-.24	***	.985	
23	Herramientas (A carpenter's tool)	7.51	9.57	-.25	***	.994	.80
24	Miembros del clero (A member of the clergy)	7.23	13.25	-.24	**	.997	
25	Condimentos (A substance for flavoring food)	8.49	16.62	-.30	***	.995	.88
26	Combustibles (A type of fuel)	5.49	12.37	-.35	***	.990	
27	Profesiones (An occupation or profession)	12.05	8.43	-.12	*	.987	.81
28	Accidentes geográficos y formaciones naturales de la tierra (A natural earth formation)	8.40	8.15	-.15	*	.987	.92
29	Deportes (A sport)	10.58	15.66	-.30	***	.996	.77
30	Fenómenos atmosféricos (A weather phenomenon)	8.56	14.22	-.29	***	.996	.90
31	Prendas de vestir (An article of clothing)	13.08	17.87	-.24	***	.991	.90
32	Partes de un edificio (A part of a building)	9.50	9.20	-.22	***	.986	
33	Elementos químicos (A chemical element)	9.51	17.53	-.02	ns	.989	
34	Instrumentos musicales (A musical instrument)	11.84	24.91	-.42	***	.996	.92
35	Dinero (A type of money)	8.60	11.57	-.29	***	.990	.77
36	Tipos de música (A type of music)	9.55	11.54	-.27	***	.994	
37	Aves (A bird)	9.25	15.02	-.20	**	.992	.92
38	Bebidas no alcohólicas (A nonalcoholic beverage)	8.71	13.16	-.26	***	.993	
39	Vehículos (A type of vehicle)	9.30	13.75	-.20	**	.996	.87
40	Ciencias (A science)	7.62	9.25	-.21	**	.992	
41	Juguetes (A toy)	8.64	6.74	-.17	***	.976	.70
42	Bailes (A type of dance)	8.10	12.71	-.25	***	.994	
43	Verduras (A vegetable)	8.27	19.92	-.27	**	.990	.90
44	Tipos de calzado (A type of footgear)	7.76	11.13	-.30	***	.991	.82
45	Insectos (An insect)	7.96	17.94	-.27	**	.991	.94
46	Nombres de mujer (A girl's first name)	16.83	8.98	-.17	***	.987	

Table 1 (continued)

#	Categories	Mean Responses	Dispersion	Correlation		Reliability	Hellinger Affinity
47	Nombres de varón (A male's first name)	15.57	11.28	-.17	***	.987	
48	Flores (A flower)	8.04	13.04	-.29	***	.994	.93
49	Enfermedades (A disease)	9.12	9.35	-.25	***	.993	.74
50	Árboles (A tree)	9.49	16.14	-.13	ns	.990	.93
51	Embarcaciones (A type of ship)	6.69	10.98	-.10	ns	.992	.81
52	Peces (A fish)	8.73	11.32	-.15	*	.990	.90
53	Serpientes (A snake)	4.31	10.63	-.22	*	.994	
54	Ciudades (A city)	14.48	18.11	-.11	ns	.988	
55	Provincias (A state)	13.63	37.94	-.16	ns	.986	
56	Universidades (A college or university)	8.35	7.67	-.13	*	.983	
	AVERAGE	**9.39**	**15.04**	**-.16**	***	**.992**	**.87**

*Note.*The numerals in the first column (#) correspond to the number of the category in Battig and Montague (1969)
ns=$p>.05$; * $p\leq.05$; ** $p\leq.01$, and *** $p\leq.001$

produced the word *j* at that position in the list; *n* is the lowest position occupied by word *j* in any list produced for the category; and *e* is a constant with the value 2.718181818459045. Following Izura and colleagues (Hernández-Muñoz et al., 2006; Izura et al., 2005), the final score was multiplied by 100 for easier understanding.

Finally, partial frequency was calculated. This value indicates the number of participants who produced the exemplar in a particular output position. The whole dataset is available for downloading as an electronic archive from the supplemental material section. This archive contains three sheets.

The first sheet (named CATEGORIES) includes the category names in Spanish, their English translation, and their corresponding number in the Battig and Montague (1969) norms. The second sheet, named AGGREGATED, displays normative information in an aggregated-response format, following a well established use in the field. In particular, singular and plural forms of what is essentially the same entity were combined into a single response, the one that was most frequently produced by the participants (e.g., in the category *Aves* (Birds), "canario" and "canarios" were combined into the single response "canario"). As shown in the top matrix in Table 2, this sheet contains the following information, arranged in columns: (1) the number of the category in the original Battig and Montague (1969) norms (CATEGORY_ID); (2) the name of the category in Spanish (English translation appears in brackets, CATEGORY_NAME); (3) the name in Spanish of the representative exemplar, combining singular and plural forms (EXEMPLAR); (4) the frequency of production of the exemplar, combining singular and plural forms (FREQUENCY); (5) the rank of each exemplar (RANK); (6) the lexical availability of each exemplar (LEXICAL AVAILABILITY), and (7) the partial frequency of the response at each of the observed output positions (ranging from 1 to 31; P1, P2, P3,…,P31). In the file, categories are ordered following the Battig and Montague (1969) sequence, and within each category, exemplars are ordered by production frequency, with most frequently produced exemplars first. The third sheet (named EXHAUSTIVE) is organized in the same way, except that, as shown in the bottom matrix in Table 2, both singular and plural responses are presented separately to provide an exhaustive list of all the exemplars produced, together with their corresponding segregated indicators.

A summary of statistics computed for each category list is shown in Table 1. The mean number of responses to each category, averaged across participants, was 9.39 (SD=2.75), with a range varying from a mean score of 4.31 in the category *serpientes* (*a snake*) to a mean score of 16.83 in the category *nombres de mujer* (*a girl's first name*). For each category, a dispersion index was obtained dividing the total number of responses by the number of unique exemplars, with higher dispersion values indicating less variability in the responses. The minimum and maximum dispersion indices were 6.29 for the category *delitos* (*a crime*) and 37.94 for the category *provincias* (*A province*), respectively. Additionally, in order to investigate the extent to which the most frequent exemplars tended to be produced in the first positions, Pearson correlation coefficients (r) between the total frequency and the rank of each exemplar were computed. All categories except five (*chemical elements, cities, four-footed animal, types of ships,* and *provinces*) showed a significant negative correlation ($p<.05$). The global correlation was−.16 ($p<.001$), implying that most frequent exemplars tended to be produced in the first output positions.

The category norms were evaluated employing the split-half method (see e.g., Ruts et al., 2004) to test the reliability of the provided responses during the exemplar generation task.

Table 2 The Organization of a Category and its Exemplars in the Normative Database

Category_ID	Category_name	Exemplar	Frequency	Rank	Lexical availability	P1	P2	[…]	P31
The Aggregated Version in the AGGREGATED sheet									
1	Piedras preciosas (A precious stone)	DIAMANTE	260	2,68	55,07	66	67	[…]	0
1	Piedras preciosas (A precious stone)	RUBÍ	258	1,95	65,51	120	68	[…]	0
1	Piedras preciosas (A precious stone)	ESMERALDA	233	3,04	49,20	26	66	[…]	0
1	Piedras preciosas (A precious stone)	ZAFIRO	110	3,90	16,79	6	14	[…]	0
1	Piedras preciosas (A precious stone)	AMATISTA	69	0,22	13,23	15	8	[…]	0
1	Piedras preciosas (A precious stone)	PERLA	51	4,39	5,75	0	8	[…]	0
1	Piedras preciosas (A precious stone)	BRILLANTE	50	4,66	6,29	2	4	[…]	0
1	Piedras preciosas (A precious stone)	TOPACIO	50	3,62	7,78	6	7	[…]	0
1	Piedras preciosas (A precious stone)	CIRCONITA	45	4,36	5,76	0	4	[…]	0
The exhaustive version in the EXHAUSTIVE sheet									
1	Piedras preciosas (A precious stone)	DIAMANTE	253	2,68	53,65	65	64	[…]	0
1	Piedras preciosas (A precious stone)	RUBÍ	246	1,90	63,38	118	65	[…]	0
1	Piedras preciosas (A precious stone) […]	ESMERALDA	228	3,03	48,36	26	66	[…]	0
1	Piedras preciosas (A precious stone) […]	RUBÍES	12	2,83	1,88	2	3	[…]	0
1	Piedras preciosas (A precious stone) […]	DIAMANTES	7	2,71	1,16	1	3	[…]	0
1	Piedras preciosas (A precious stone)	ESMERALDAS	5	3,6	0,26	0	0	[…]	0

CATEGORY_ID: category number in Battig and Montague (1969); CATEGORY_NAME: name of the category in Spanish (English translation in brackets); EXEMPLAR: name of the produced exemplar in Spanish; FREQUENCY: number of participants producing the exemplar, out of 284; RANK: Mean output position of the response; LEXICAL AVAILABILITY: ease with which a word is produced as a member of one category; P1 to P31: Partial frequencies, i.e., number of participants producing the exemplar in each output position

Thus, participants were randomly divided into two equal groups and the resulting frequencies in each group were correlated employing the Spearman-Brown correction. Results showed correlation estimates very close to 1 (the lowest value was .97), an indication of a high reliability (see the reliability values in the last column of Table 1).

The present norms were also contrasted with the ones previously reported by Soto et al. (1982). An inspection of the most frequent exemplar in each of the 32 common categories showed an exact match in 66 % of the cases, indicating a certain degree of stability over time. However, we also observed some differences with the passage of time. Unsurprisingly, the most frequent exemplar for the category *dinero* (*money*) in the present study was *euro*, while, in Soto et al., it was the no longer used *peseta*. In addition, the frequency of some exemplars had also changed: for example, idiosyncratic or non-evoked items in Soto et al. are relatively frequent exemplars in the present study (e.g. *kiwi, mango, papaya* in the category *fruit*), reflecting changes in availability and use over time.

Taking into account both similarities and discrepancies between these two studies we computed a numerical estimate to check the generational stability across norms and, additionally, the validity of the results obtained in the present study. Traditionally, correlational analyses have been carried out to make this comparison (e.g., Van Overschelde et al., 2004). However, as Yoon et al. (2004) have explained in detail, correlations can be misleading when comparing category norms. First, the correlation depends on how many items are included in the analysis. For example, the most frequent items can be inversely correlated (i.e., correlation estimate is near −1) in the two category norms. Adding a large number of idiosyncratic exemplars to the sample, however, can change this relation dramatically into the positive direction. A second problem appears when there is zero overlap between categories, i.e., when two categories have no exemplars in common the correlation will be zero. A correlation close to zero indicates that there is no relation between the exemplar frequencies, but this value does not reflect the lack of overlap between category exemplars. Finally, the weight of each exemplar in the correlational analysis is the same, independently of its higher or lower frequency. To solve these problems, the use of an index called Hellinger Affinity (HA) has been proposed (Yoon et al., 2004). HA assigns more weight to more frequent exemplars, does not depend on the amount of items included in the analysis, and, importantly, indicates the degree of overlap between category exemplars.

Given two categories with n exemplars, and given the two frequency distributions for their exemplars (p_1, p_2, p_3,…, p_n) and (q_1, q_2, q_3,…,q_n), HA is calculated by summing the square root of the product of the two exemplar frequencies.

$$HA(p,q) = \sum_{i=1}^{n} \sqrt{p_i q_i}$$

HA values range from 1 to 0. An HA value of 1 corresponds to two identical distributions; an HA of 0 indicates no overlap between the exemplar frequencies.

HA values resulting from a comparison between the present study and Soto et al. (1982) are presented in Table 1. The mean HA was .87 (SD=.07), indicating an overall high overlap between the two category norms. An inspection of the indices revealed that the category *parientes* (*relatives*) had the highest HA value (.96), indicating a high overlap between norms. The categories *metales* (*metals*) and *unidades de tiempo* (*units of time*) also showed high HA values (.95 in both cases). On the other hand, the category *juguetes* (*toys*) showed the lowest HA value (.70). For instance, the popular doll *Barbie* was a frequent exemplar in the present norms (75 responses), but it was not produced at all in Soto et al. Finally, the category *enfermedades* (*diseases*) also had a relatively low HA value (.74). In this category the disease *SIDA* (*AIDS*), was the second most frequent exemplar in the present study (209 responses), but it was not elected by any participant in the Soto et al. norms.

Discussion

This study collected production data for exemplars in the 56 categories of Battig and Montague (1969) from a pool of 284 young Spanish adults. Frequency, rank, lexical availability, and partial frequency indices were obtained for each exemplar.

As described earlier, the norms have high reliability and appropriate validity, and comparative analyses showed that they tend to be stable over time, with a few exceptions that underscore the need for users of the normative data to be aware of changing context and conditions of use over time. On a related point, it should be noticed that there are likely to be geographic constraints in the organization of lexical and semantic knowledge that might limit the application of any kind of word norms. Therefore, direct utilization of the norms provided here by other Spanish-speaking populations (e.g., native speakers in Argentina, Colombia, Mexico, the USA, etc.) must be made with caution.

Nonetheless, these data are expected to be a valuable resource in different ways. Indices such as exemplar frequencies or ranks are used habitually in psychological experiments using verbal material, as research frequently requires the control of a wide range of variables (e.g., frequency, typicality, associative strength, age of acquisition, and others, as well as categorical information) for a given set of stimuli. Furthermore, categorical norms can also be useful in more specific areas, for example, crosscultural comparisons. It is worth noting that previous category norms available in Spanish did not contain all the categories of Battig and Montague (1969). The selection of the same categories used by Battig and Montague in the present study has the great advantage of allowing comparisons with most of the category norms used around the world. Categorical exemplar identification and retrieval is also crucial in neuropsychological studies of patients with brain damage or dementias. For instance, Henry, Crawford, and Phillips (2004) conducted a meta-analysis of 153 studies in patients with Alzheimer type dementia. In total, 126 publications contained measures of semantic fluency (e.g., generation of exemplars from the category animals). The use of category-based indicators of semantic status in dementias is also evident in recent research with Spanish-speaking participants (e.g., Rodríguez-Ferreiro et al. 2012). Finally, category norms are also useful instruments in more applied settings. Thus, as Hernández-Muñoz et al. (2006) explained, categorical data can be used for establishing empirically based criteria when selecting specific materials to develop concise dictionaries, educational tools for primary education, and learning materials for second-language acquisition. The availability of the norms presented here is likely to make a helpful contribution to the design of better studies and better applied tools in the near future.

Finally, it is worth noting that some normative studies in Spanish have aimed, in the past, at collecting various relevant indices for a particular set of materials. For example, empirical work by Izura and colleagues (Hernández-Muñoz et al., 2006; Izura et al., 2005) has provided categorical information plus valuable indices such us lexical availability, typicality, and age of acquisition, among others. This strategy produces a relatively wide range of lexical characteristics for the items of interest, but it has the limitation of being focused on a limited set of words. An alternative, more recent approach to increase the amount of information about words is the utilization of data obtained in different studies by different investigators. For instance, the Spanish database NIPE (Norms and Indices for Experimental Psychology; Díez, Fernandez, & Alonso, 2014) contains information from many Spanish normative studies published in scientific papers and facilitates online access to quantitative information on a wide range of psycholinguistic variables for user-specified sets of words. In addition, recently constructed databases such as EsPal (Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013) or NIM (Guasch, Boada, Ferré, & Sánchez-Casas, 2013) offer consultation on a set of psycholinguistic indices (e.g., written frequency, orthographic neighbors, number of letters, etc.) for a

large body of words that are mainly based on written materials. It is to this latter accumulative approach that the norms presented here can be a substantial contribution, not only because of the large number of individual words, acronyms, and, compound words now having category-membership information (i.e., these norms contain 6,762 distinct individual words and acronyms, and 1,746 distinct compound words), but also because these norm words are well represented in other normative studies readily available for the Spanish language. Thus, for example, 91 % of the individual words provided in the present norms also appear in the LEXESP database (Sebastián, Martí, Carreiras, & Cuetos, 2000). In addition, age of acquisition (Alonso, Fernandez, & Díez, in press) and oral frequency (Alonso, Fernandez, & Díez, 2011) values can be obtained for 33 % and 69 % of these individual exemplars, respectively.

Author Note Alejandra Marful was supported by the Spanish Ministerio de Educación y Ciencia, through a Juan de la Cierva postdoctoral scholarship. Additional funding was provided to Emiliano Díez and Angel Fernandez by Ministerio de Ciencia e Innovación (Project PSI2008-05607) and by Universidad de Salamanca (Plan Propio de Apoyo a la Investigación 2013). We thank Fred Feinberg for his help in the calculation of Hellinger Affinity indices.

References

Abel, M., & Bauml, K.-H. T. (2013). Sleep can eliminate list-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 946–952.

Alonso, M. A., Fernandez, A., & Díez, E. (2011). Oral frequency norms for 67,979 Spanish words. *Behavior Research Methods, 43,* 449–458.

Alonso, M. A., Fernandez, A., & Díez, E. (in press). Subjective Age-of-Acquisition norms for 7,039 Spanish words. *Behavior Research Methods*.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063–1087.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 629–654.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80,* 1–46.

Boccardi, M., & Cappa, S. F. (1997). Valori normative di produzione categoriale per la lingua italiana [Normative values of categorical production for the Italian language]. *Giornale Italiano di Psicologia, 24,* 425–436.

Brown, W. P., & Davies, G. M. (1976). Studies in word listing: Testing for group differences in category norms. *Irish Journal of Psychology, 3,* 87–120.

Bueno, S., & Megherbi, H. (2009). French categorization norms for 70 semantic categories and comparison with Van Overschelde et al.'s (2004) English norms. *Behavior Research Methods, 41,* 1018–1028.

Carneiro, P., Albuquerque, P., & Fernandez, A. (2008). Portuguese category norms for children. *Behavior Research Methods, 40,* 177–182.

Carneiro, P., Albuquerque, P., & Fernandez, A. (2009). Opposite developmental trends for false recognition of basic and superordinate names. *Memory, 17,* 411–427.

Casey, P. J., & Heath, R. A. (1988). Category norms for Australians. *Australian Journal of Psychology, 40,* 323–339.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. (1957). *Cultural norms for verbal items in 43 categories*. Storrs, CT: University of Connecticut.

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers, 31,* 650–658.

Díez, E., Fernandez, A., & Alonso, M. A. (2014). NIPE: Normas e índices de interés en Psicología Experimental. Accesible at http://campus.usal.es/~gimc/nipe/

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods, 45,* 1246–1258.

Goikoetxea, E. (2000). Frecuencia de producción de las respuestas a 52 categorías verbales en niños de primaria [Frequency of response production to 52 verbal categories in school children]. *Psicológica, 21,* 61–89.

Guasch, M., Boada, R., Ferré, P., & Sánchez-Casas, R. (2013). NIM: A Web-based Swiss army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods, 45,* 765–771.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia, 42,* 1212–1222.

Hernández-Muñoz, N., Izura, C., & Ellis, A. W. (2006). Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology, 18,* 730–755.

Howard, D. V. (1980). Category norms: A comparison of Battig and Montague (1969) norms with the responses of adults between the ages of 20 and 80. *Journal of Gerontology, 35,* 225–231.

Izura, C., Hernández-Muñoz, N., & Ellis, A. W. (2005). Category norms for 500 Spanish words in five semantic categories. *Behavior Research Methods, 37,* 385–397.

Jenkins, J. J., & Palermo, D. S. (1965). Further data on changes in word association norms. *Journal of Personality and Social Psychology, 1,* 303–309.

López-Chávez, J., & Strassburguer-Frías, C. (2000). El diseño de una fórmula matemática para obtener un índice de disponibilidad léxica confiable [A mathematical equation designed to obtain a reliable lexical availability index]. *Anuario de Letras, 38,* 227–251.

Macizo, P., Bajo, M. T., & Paolieri, D. (2012). Language switching and language competition. *Second Language Research, 28,* 131–149.

Marshall, C. E., & Parr, W. V. (1996). New Zealand norms for a subset of Battig and Montague's (1969) categories. *New Zealand Journal of Psychology, 25,* 24–29.

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS ONE, 7*(5), e37527.

Pascual, J., & Musitu, G. (1980). Normas categoriales [Categorical norms]. *Psicológica, 1,* 157–174.

Pinto, A. C. (1992). Medidas de categorização: Frequência de produção e de tipicidade [Measures of categorization: Production frequency and typicality]. *Jornal de Psicologia, 10,* 10–15.

Price, H. L., & Connolly, D. A. (2006). BatMon II: Children's category norms for 33 categories. *Behavior Research Methods, 38,* 529–531.

Rodríguez-Ferreiro, J., Cuetos, F., Monsalve, A., Martínez, C., Pérez, A. J., & Venneri, A. E. (2012). Establishing the relationship between cortical atrophy and semantic deficits in Alzheimer's disease and mild cognitive impairment patients through voxel-based morphometry. *Journal of Neurolinguistics, 25,* 139–149.

Ruts, W., De Deyne, S., Ameel, E., Vanpaemel, W., Verbeemen, T., & Storms, G. (2004). Dutch norm data for 13 semantic categories and

338 exemplars. *Behavior Research Methods, Instruments, & Computers, 36,* 506–515.

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass-Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers, 28,* 537–555.

Saunders, J., Fernandes, M., & Kosnes, L. (2009). Retrieval-induced forgetting and mental imagery. *Memory & Cognition, 37,* 819–828.

Sebastián, N., Martí, M. A., Carreiras, M. F., & Cuetos, F. (2000). *LEXESP, léxico informatizado del español*. Barcelona, Spain: Ediciones de la Universitat de Barcelona.

Smith, S. M., Ward, T. B., Tindell, D. R., Sifonis, C. M., & Wilkenfeld, M. J. (2000). Effects of category structure on created memories. *Memory & Cognition, 28,* 386–395.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Soto, P., Sebastián, M. V., García, E., & del Amo, T. (1982). *Categorización y datos normativos en España [Categorization and normative data in Spain]*. Madrid: Universidad Autónoma de Madrid.

Soto, P., Sebastián, M. V., García, E., & del Amo, T. (1994). *Las categorías y sus normas en castellano [Categories and their norms in Spanish]*. Madrid: Visor.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289–335.

Yoon, C., Feinberg, F., Hu, P., Gutchess, A. H., Hedden, T., Chen, H. Y. M., ... Park, D. C. (2004). Category norms as a function of culture and age: Comparisons of item responses to 105 categories by American and Chinese adults. *Psychology and Aging, 19,* 379–393.

Zacks, R. T., Radvansky, G., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 143–156.