



Eliciting semantic properties: methods and applications

Sergio E. Chaigneau¹ · Enrique Canessa^{1,2} · Alessandro Lenci³ · Barry Devereux⁴

Received: 1 October 2020 / Accepted: 5 October 2020

© Marta Olivetti Belardinelli and Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Asking subjects to list semantic properties for concepts is essential for predicting performance in several linguistic and non-linguistic tasks and for creating carefully controlled stimuli for experiments. The property elicitation task and the ensuing norms are widely used across the field, to investigate the organization of semantic memory and design computational models thereof. The contributions of the current Special Topic discuss several core issues concerning how semantic property norms are constructed and how they may be used for research aiming at understanding cognitive processing.

The property listing task (PLT, a.k.a. Feature Listing Task) and the semantic property norms (SPNs, a.k.a. Semantic Feature Norms) derived from this task are widely used by researchers interested in cognition. Asking subjects to list semantic properties for concepts has proven very useful for creating carefully controlled stimuli for experiments and for predicting performance in many linguistic and non-linguistic tasks. The property elicitation task and the ensuing norms are widely used to investigate the organization of semantic memory and to design computational models of concept representation and access. The contributions to the current Special Topic discuss several core issues concerning how SPNs are constructed and how they may be used for research aiming at understanding cognitive processing. In doing so, the Special Topic illustrates the breadth of applications of SPNs and contributes to standardizing procedures that could in the future enable better large-scale studies (i.e., inter-laboratory, inter-language, and inter-cultural comparisons),

thus expanding and developing the use of norming studies in cognition and related fields.

In general, researchers using the PLT wish to characterize conceptual content coming from semantic memory (Canessa and Chaigneau 2020; Chaigneau et al. 2018). In a PLT, participants are asked to list properties that are typically true of a given concept. After the PLT is used to extract semantic properties for individual concepts, SPNs can then be constructed from the elicited features to characterize the representation and content of a semantic space. SPNs are generally represented as matrices containing different concepts with their respective properties' frequency distributions (e.g., Devereux et al. 2014; Kremer and Baroni 2011; Lenci et al. 2013; McRae et al. 2005; Montefinese et al. 2013; Vivas et al. 2017).

The PLT and associated SPNs have been used in basic cognitive research, as well as in applied or field studies (in cognitive psychology, e.g., Wu and Barsalou 2009; in marketing, e.g., Hough and Ferraris 2010; in social psychology, e.g., Walker and Hennig 2004; in neuropsychology, e.g., Perri et al. 2012). Importantly, these norms are widely used as a way to test theories (e.g., Cree and McRae 2003; Vigliocco et al. 2004; Wu and Barsalou 2009; Taylor et al. 2011), and as a source of carefully controlled experimental stimuli (e.g., McRae et al. 1999; Bruffaerts et al. 2019). SPNs have attracted attention in computational linguistics too, as benchmarks to evaluate the possibility of extracting conceptual knowledge from corpora (Baroni and Lenci 2008; Devereux et al. 2009; Făgărășan et al. 2015).

Notwithstanding their wide use and relevance, the PLT and the development of the ensuing SPNs still face several challenges (Bolognesi et al. 2016; Canessa et al. 2020).

✉ Sergio E. Chaigneau
sergio.chaigneau@uai.cl

¹ Center of Cognitive Research (CINCO), Universidad Adolfo Ibáñez, Avda. Presidente Errázuriz, 3328 Las Condes, Santiago, Chile

² Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Viña del Mar, Chile

³ Dipartimento Di Filologia, Letteratura E Linguistica, Università Di Pisa, Pisa, Italy

⁴ School of Electronics, Electrical Engineering and Computer Science, Queens University Belfast, Belfast, Northern Ireland, UK

Surprisingly, the procedures that researchers follow when developing SPN studies are based on an implicit consensus regarding how these kinds of studies should be carried out. There are many decisions that researchers have to make, often with little explicit discussion regarding what each of these decisions imply for the semantic model. Relatedly, making inferences about semantic structure based on SPN data continues to be somewhat of an art. Under these conditions, performing and interpreting large-scale studies is complicated, as is making comparisons across data collected by different laboratories, in different cultures, and across different languages. The contributions to this Special Topic confront some of these issues head-on, offer solutions, and give examples of the use of SPNs to study semantic memory.

Regarding the need for more explicit design criteria for SPN studies, in “A practical primer on processing semantic property norm data,” Buchanan et al. (2020) provide several recommendations on how to collect and process semantic property norms. The work discusses methods to elicit semantic properties and how to derive meaningful representations. It also discusses decisions that need to be made at different stages of an SPN study: instructions and test context, property preprocessing, property weighting, and relationship encoding using ontologies. By explicitly considering each of these steps, the paper offers a processing pipeline that could result in new property listing norms that overcome the constraints of previous studies, and allow improved comparability across SPNs.

Having presented a possible standardization for carrying out SPNs, the next three papers provide examples of how to use SPN data to make inferences about semantic structure. In “Informational content of cosine and other similarities calculated from high-dimensional Conceptual Property Norm data,” Canessa et al. (2020) discuss several inter-concept similarity measures used in characterizing semantic spaces, including the cosine of conceptual property frequency vectors. This last similarity measure has become a de facto standard without much scrutiny. To provide a comparative study that tests the merit of different similarity measures when computed from property frequency data, this paper compares four different similarity measures and five different types of data structures. Similarity distributions were compared regarding their informational content (i.e., entropy), and a clustering procedure was used as a concrete example of how informational content affects statistical analyses. In general, the results show that similarity measures computed from lower-dimensional data offer more information than higher-dimensional data, such as those resulting from cosine similarity. Typically, SPN data are sparse and high-dimensional, and this fact has important consequences for statistical analyses.

Also in the context of extracting semantic information from SPN studies, in “Emoji-based semantic representations

for abstract and concrete concepts” Wicke and Bolognesi (2020) explore a different method for collecting conceptual properties. Typically, such properties are collected from participants as verbal strings. However, if the semantic system is distributed and multimodal, a theory of semantic representations should take into account different modalities in which property-based representations are generated, some of which may not be easily transformed into verbal productions that participants can report. Accordingly, the authors asked participants to use emoji to provide semantic representations for a sample of 300 English nouns referring to abstract and concrete concepts. Using a content analysis procedure with multiple annotators, the authors classified the cognitive strategies used by the participants to represent conceptual content through emoji. Interestingly, this procedure produced results globally consistent with prior research in that abstract concepts were found to produce a larger number of emoji, include more face emoji expressing emotions, and be more variable across users.

Related to this same theme of extracting semantic information from SPNs, semantic richness is a variable that has generated much research interest. Perhaps the simplest way to measure this construct is to count the number of properties or features associated with a given concept (NoF), variable which is easily obtained from SPNs. Several studies have shown NoF to be an important variable that affects cognitive processing. However, those studies relied on a small set of words. To overcome this limitation, in the paper “Mapping semantic space: property norms and semantic richness,” Muraki et al. (2020) test for semantic richness effects using a larger item set. Additionally, the authors also explore how NoF relates to other measures of semantic richness, including subjective ratings of different kinds, as well as more objective measures like semantic diversity, number of associates, and lexical centrality. Related to discussions in the previous paper, results are interpreted as evidence that semantic representations are multimodal and multidimensional, providing insights about the structure of semantic space.

Finally, the next two papers illustrate the use of SPNs in inter-study comparisons and large-scale studies. As previously discussed, we envision that being able to make inter-study comparisons and to engage in large studies, perhaps of a collaborative nature, should be a long-term goal of researchers engaged in SPN development. In this spirit, in “Core features: measures and characterization for different languages,” Vivas et al. (2020) compare property informativeness and distinctiveness in data collected in SPNs across different languages. This paper studies language-related differences and similarities emerging from the semantic core features, leading to a characterization of concrete concepts in terms of their feature types. Interestingly, results show that the characterization of concrete concepts is similar across

languages, which leads the authors to reflect on the theoretical paradigm of semantic property norms.

To close, the Special Topic concludes with the study “Feature distinctiveness in language acquisition and lexical processing: insights from megastudies,” which illustrates the use of megastudies to examine the influence of property distinctiveness on language acquisition, visual lexical decision, and semantic decision (Siew 2020). By representing semantic relations as a network where edges represent dissimilarities between words, property distinctiveness for individual words was quantified. Results showed that property distinctiveness had effects on several dependent measures, thus demonstrating the importance of considering words as embedded in a semantic space as a means of understanding their processing and acquisition.

The diversity of approaches covered in the Special Topic papers shows the breadth of possible applications and the usefulness of the PLT and SPNs to advance our theories and understanding of the conceptual system. It also exposes many of the open issues and problems still not completely solved when collecting and analyzing SPN data, all of which have important consequences when carrying out such studies. Thus, we are optimistic that this Special Topic will be a stepping stone toward solving some of these problems and broadening the research community’s understanding of these techniques and their range of applications.

Funding This study was funded by an ANID, Fondo Nacional de Desarrollo Científico y Tecnológico, Grant 1200139 to the second author.

Compliance with ethical standards

Conflict of interest Sergio E. Chaigneau declares he has no conflict of interest. Enrique Canessa declares he has no conflict of interest. Barry Devereux declares he has no conflict of interest. Alessandro Lenci declares he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

- Baroni M, Lenci A (2008) Concepts and properties in word spaces. From Context Mean Distrib Models Lex Linguist Cogni Sci (Spec Issue Ital J Linguist) 20(1):55–88
- Bolognesi M, Pilgram R, van den Heerik R (2016) Reliability in content analysis: the case of semantic feature norms classification. *Behav Res Methods* 49(6):1984–2001. <https://doi.org/10.3758/s13428-016-0838-6>
- Buchanan EM, De Deyne S, Montefinese M (2020) A practical primer on processing semantic property norm data. *Cogn Process*. <https://doi.org/10.1007/s10339-019-00939-6>
- Bruffaerts R, De Deyne S, Meersmans K, Liuzzi AG, Storms G, Vandenberghe R (2019) Redefining the resolution of semantic knowledge in the brain: advances made by the introduction of models of semantics in neuroimaging. *Neurosci Biobehav Rev* 103:3–13. <https://doi.org/10.1016/j.neubiorev.2019.05.015>
- Canessa E, Chaigneau SE, Moreno S, Lagos R (2020) Informational content of cosine and other similarities calculated from high-dimensional conceptual property norm data. *Cogn Process*. <https://doi.org/10.1007/s10339-020-00985-5>
- Canessa E, Chaigneau SE (2020) Mathematical regularities of data from the property listing task. *J Math Psychol*. <https://doi.org/10.1016/j.jmp.2020.102376>
- Canessa E, Chaigneau SE, Lagos R, Medina FA (2020) How to carry out conceptual properties norming studies as parameter estimation studies: lessons from ecology. *Behav Res Methods*. <https://doi.org/10.3758/s13428-020-01439-8>
- Chaigneau SE, Canessa E, Barra C, Lagos R (2018) The role of variability in the property listing task. *Behav Res Methods* 50(3):972–988. <https://doi.org/10.3758/s13428-017-0920-8>
- Cree GS, McRae K (2003) Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *J Exp Psychol Gen* 132(2):163–201. <https://doi.org/10.1037/0096-3445.132.2.163>
- Devereux BJ, Tyler LK, Geertzen J, Randall B (2014) The centre for speech, language and the brain (CSLB) concept property norms. *Behav Res Methods* 46(4):1119–1127. <https://doi.org/10.3758/s13428-013-0420-4>
- Devereux BJ, Pilkington N, Poibeau T, Korhonen A (2009) Towards unrestricted, large-scale acquisition of feature-based conceptual representations from corpus data. *Res Lang Comput* 7(2–4):137–170
- Făgărășan L, Vecchi EM, Clark S (2015). From distributional semantics to feature norms : grounding semantic models in human perceptual data. In: Proceedings of the 11th international conference on computational semantics, London, UK, 15–17 April 2015, p 52–57
- Hough G, Ferraris D (2010) Free listing: a method to gain initial insight of a food category. *Food Qual Prefer* 21(3):295–301. <https://doi.org/10.1016/j.foodqual.2009.04.001>
- Kremer G, Baroni M (2011) A set of semantic norms for German and Italian. *Behav Res Methods* 43(1):97–109. <https://doi.org/10.3758/s13428-010-0028-x>
- Lenci A, Baroni M, Cazzolli G, Marotta G (2013) BLIND: a set of semantic feature norms from the congenitally blind. *Behav Res Methods* 45(4):1218–1233. <https://doi.org/10.3758/s13428-013-0323-4>
- McRae K, Cree GS, Seidenberg MS, Mcnorgan C (2005) Semantic feature production norms for a large set of living and nonliving things. *Behav Res Methods* 37(4):547–559. <https://doi.org/10.3758/BF03192726>
- McRae K, Cree GS, Westmacott R, de Sa VRD (1999) Further evidence for feature correlations in semantic memory. *Can J Exp Psychol/Rev can psychol exp* 53(4):360–373. <https://doi.org/10.1037/h0087323>
- Montefinese M, Ambrosini E, Fairfield B, Mammarella N (2013) Semantic memory: a feature-based analysis and new norms for Italian. *Behav Res Methods* 45(2):440–461. <https://doi.org/10.3758/s13428-012-0263-4>
- Muraki EJ, Sidhu DM, Pexman PM (2020) Mapping semantic space: property norms and semantic richness. *Cogn Process*. <https://doi.org/10.1007/s10339-019-00933-y>
- Perri R, Zannino G, Caltagirone C, Carlesimo GA (2012) Alzheimer’s disease and semantic deficits: a feature-listing study. *Neuropsychology* 26(5):652–663. <https://doi.org/10.1037/a0029302>

- Siew CS (2020) Feature distinctiveness effects in language acquisition and lexical processing: insights from megastudies. *Cogn Process.* <https://doi.org/10.1007/s10339-019-00947-6>
- Taylor KI, Devereux BJ, Tyler LK (2011) Conceptual structure: towards an integrated neurocognitive account. *Lang Cogn Process* 26(9):1368–1401
- Vigliocco G, Vinson DP, Lewis W, Garrett MF (2004) Representing the meanings of object and action words: the featural and unitary semantic space hypothesis. *Cogn Psychol* 48(4):422–488. <https://doi.org/10.1016/j.cogpsych.2003.09.001>
- Vivas J, Vivas L, Comesaña A, García Coni A, Vorano A (2017) Spanish semantic feature production norms for 400 concrete concepts. *Behav Res Methods* 49(3):1095–1106. <https://doi.org/10.3758/s13428-016-0777-2>
- Vivas L, Montefinese M, Bolognesi M, Vivas J (2020) Core features: measures and characterization for different languages. *Cogn Process.* <https://doi.org/10.1007/s10339-020-00969-5>
- Walker LJ, Hennig KH (2004) Differing conceptions of moral exemplarity: just, brave, and caring. *J Pers Soc Psychol* 86(4):629–647. <https://doi.org/10.1037/0022-3514.86.4.629>
- Wicke P, Bolognesi M (2020) Emoji-based semantic representations for abstract and concrete concepts. *Cogn Process.* <https://doi.org/10.1007/s10339-020-00971-x>
- Wu LL, Barsalou LW (2009) Perceptual simulation in conceptual combination: evidence from property generation. *Acta Physiol (Oxf)* 132:173–189. <https://doi.org/10.1016/j.actpsy.2009.02.002>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.