***International Encyclopedia of Language and Linguistics, Third Edition***

**Article Title**

Semantic Feature Production Norms

**Author and Co-author Contact Information**

Erin M. Buchanan

326 Market St., Harrisburg, PA, USA 17101

ebuchanan@harrisburgu.edu

+1 (717) 901-5100

**Abstract**

Feature production norms are collected by asking participants to name the defining features of a concept, such as *tail, fur,* and *meow* for the concept *cat*. Collection and analysis of features should include special considerations for methodology and data processing including the task instructions, segmentation, word removal, spell checking and more. The processed data can be used to define concept similarity, control stimuli for new experimental studies, and in computational models of memory and knowledge representation.

**Keywords**

Datasets, features, knowledge, memory, norms, production task, semantics

**Key points**

* The feature production task is used to collect the feature production norms, which consists of asking participants to define key features of individual concepts.
* This task is similar to, but provides different information from, the free association task in which participants are asked to list the first word that comes to mind when a concept is presented.
* When collecting feature production norms, special care should be given to thinking about how the instructions are presented to participants and how the data should be processed.
* The feature production norms are a useful tool for creating new experimental stimuli and answering new research questions related to another linguistic variables.
* The feature production norms have been used in computational modelling to create semantic networks of memory and knowledge representation, which were an early precursor to large language models.

**Introduction**

Humans have a rich semantic memory that creates a dictionary of knowledge from which we interpret and explain the world. This memory contains information about concepts, their meanings, the relationship between concepts, and other facts (Tulving, 1972). This information is the cornerstone to everyday existence, allowing people to discriminate between objects, even when those objects are very similar or contain many variations. For example, imagine a *dog*. There are many types of dogs, ranging from small (Chihuahuas) to very large (Great Danes), and they come in many colors, shapes, and temperaments. Yet, it is easy for people to know a dog when they see one and to distinguish dogs from another common pet, *cat*. To understand human object recognition, we first have to understand what features people believe an object possesses, that is, what features are salient (i.e., most noticeable or important) when describing an object. In this chapter, we will discuss the feature production task, the corresponding data created from this task, and the importance of large datasets to the study of linguistics, psychology, and computational fields.

**Feature production task**

The simplest way to access information about the features that define a particular object is to ask people using the feature production task. In the feature production task, participants are required to list the features or properties of an object that “make it” that concept. For example, here are the instructions given to participants in Vinson and Vigliocco (2008):

*In this experiment, you will be asked to produce definitions for common English words. However, instead of writing "dictionary style" definitions, we want you to define the words using features (described below).*

*Each feature should contain as few words as possible. The features you list for a word, when they are combined, should define and describe that particular word as completely as possible. Think about the features of meaning that are most important for each word and try to list features that will uniquely identify that word even among similar words.*

Participants are then given two examples: a noun (*dog*) and a verb (*to write*). Dog features included *pet, animal, has fur, barks*, and writing features included *communication, action, requires paper,* and *uses words*. Participants are then given a list of concepts to define with general preference toward concrete nouns (*car*, Devereux et al., 2014; McRae et al., 2005) in the research, with updated research on verbs (*drive*, Vinson & Vigliocco, 2008), and abstract concepts (*truth*, Buchanan et al., 2019). The examples presented here are in English; however, this task has also been used in Spanish (Vivas et al., 2017), Italian (Montefinese et al., 2013), German (Kremer & Baroni, 2011), Portuguese (Stein & Gomes, 2009), and Dutch (Ruts et al., 2004) to name a few. Table 1 shows some of the common features listed with the concepts of *cat*, *mouse*, and *computer* from Buchanan et al. (2019):

Table 1. *Most Common Feature Responses to Cat, Mouse, and Computer*

|  |  |  |
| --- | --- | --- |
| Cat | Mouse | Computer |
| Fur | Rodent | Internet |
| Animal | Small | Technology |
| Small | Tail | Keyboard |
| Pet | Cheese | Electronic |
| Four Legs | Animal | Typing |
| Meow | Fur |  |
| Domestic | Gray |  |
| Tail | Squeak |  |

These most common features tell us a few things about recognition and the structure of our internal dictionary. First, the features are most commonly “is a” and “has a” types of descriptions, which mirrors early theoretical models of the structure of semantic memory (Collins & Quillian, 1969). The full list of features also includes uses, locations, behavior, and gender markers (i.e., *actress* versus *act* or *actor*). Another interesting component to these features is the overlap between two concepts that we know are related (*cat*-*mouse*), such as *fur, animal,* and *small* but also their differences such as *pet, meow,* and *squeak.* Therefore, it is conceivable we use these features to determine both that items are similar and to distinguish between them. However, we also know that *mouse*-*computer* are related, as a *mouse* is a common computer accessory. We do not see any overlap between the defining features of *mouse* and *computer* which indicates that this task does not always capture relations between objects (i.e., *computer* “uses-a” *mouse*). These large-scale sets of participant answers are then normalized (i.e., *light, lighter, and lights* all become the feature *light*) and summarized across participants to represent the salient features of concepts. As shown in Table 1, these lists are not exhaustive for features and only represent features that participants immediately report as defining. For example, *cats* have *hearts* and *skin,* but these features are unlikely to be listed in the feature production task because most animals share these features.

**Free association task**

Why don’t we see any overlap between *mouse* and *computer*? The feature production task and corresponding norms are thought to represent the dictionary component to semantic memory and how we organize knowledge and concept representations (Taylor et al., 2007). Similarity between any two concepts can be calculated by examining the overlap in features and their strength (i.e., the number of participants who mention each feature). Consider, however, the experiences of individuals that lead to understanding that *mouse* and *computer* are related through using a mouse to operate a computer. These experiences may be represented by direct associations in our mental dictionary through repeated co-occurrence in language and text (De Deyne et al., 2013). To study this component to our semantic knowledge, researchers use a free association task where participants are simply as to “list the first (several) words that come to mind”. Table 2 indicates the strongest free association responses to *cat, mouse,* and *computer* from the Small World of Words Project (De Deyne et al., 2019)*.*

Table 2. *Most Common Associative Responses to Cat, Mouse, and Computer*

|  |  |  |
| --- | --- | --- |
| Cat | Mouse | Computer |
| Dog | Cat | Internet |
| Feline | Cheese | Laptop |
| Meow | Computer | Keyboard |
| Mouse | Trap | Screen |
| Purr | Rat | Mac |
| Fur | Rodent | Mouse |
| Kitten | Small  | Work |
| Soft | Grey | Technology |

Given that both tasks are thought to shed light on our knowledge representation, it is not surprising that there is an overlap between the answers. The answers do not completely overlap, which indicates that the feature production task does gather different data than the free association task (as the overlap in answers is often approximately 30-40%, Buchanan et al., 2019). The results from this task show that *mouse* and *computer* are related, computer and mouse are both popular answers when asked to think of those concepts.

**Considerations for the feature production task and data**

The data provided from the feature production task is an invaluable resource (see Applications section). Yet, careful consideration should be employed when collecting, processing, analyzing, and putting those data to use. First, researchers should consider the methodology used to collect the data. The instructions of the task are designed to elicit defining features, rather than free associations, but the examples given may bias participant answers towards specific types of knowledge rather than simply elucidating the knowledge that is known. Early feature production studies were collected via pencil and paper, and later work was collected via computer and the internet. The Centre for Speech, Language and the Brain (CSLB) norms gave participants a dropdown box with *is, has, does, made of*, and an “…” (i.e., other) category to use to demarcate their responses. In the Buchanan et al. (2019) data collection, participants were simply given an open text box to list several features. Each methodology has its advantages: is-a, has-a relationships are more clearly defined, but may bias participant answers; open text responses may lead to less biased answers but more singleton answers (i.e., answers only one individual gave) and can be much more difficult to process given that people often used full sentences, rather than one or several word phrases.

Second, the process of preparing the data for reuse and analysis can be quite tedious, even with the present computerized toolkits available (Buchanan et al., 2020). The individual responses can be broken down into single tokens (i.e., words) or multi-token phrases (e.g., *four legs*), which requires different types of text segmentation processing. The words can then be lower-cased so they can be tallied; yet, it is important to remember than *apple* and *Apple* are not necessarily the same feature answer. Certain types of tokens are often removed, such as punctuation or stop words (e.g., *an, of, the, on*), with the special consideration if numbers are to be included or removed. If they are included, one should determine if they will be “spelled out” or left in numerical form, and the same process could be considered for contractions (e.g., can’t, don’t). The two most difficult components to processing feature responses include spell checking and the decision on how to combine like forms through lemmatization (i.e., converting a word into its root form). Automated spell checkers can be used for simple misspellings, but it is sometimes difficult to know exactly which word suggestion would be correct without context. Last, while it seems obvious that *run, ran, running,* and *runs* are all the same concept, one has to decide if the information from the affix or tense conversion is important to the feature definition. For example, *prince* and *princess* are unlikely to be combined, just like *husband* and *wife* are generally kept separate – should *actor* and *actress* then be left separate or combined into *act*?

Once these decisions are made, the data can be combined to calculate the number of individuals who listed each feature for a target concept. Before any analyses, the number of feature singletons should be considered. Potentially, these single responses may indicate an important, but infrequently mentioned feature of an object or they could simply represent individual experience that is not representative of a larger concept representation. The choice of how many and what percent of singletons to remove for analyses may impact results, which should be carefully considered. One of the most popular norm sets, McRae et al. (2005) used a cut-off of at least 16% (i.e., 5/30 participants) had to mention a feature for it to be included, and others have followed suit, but this scoring is an open question within feature production norm data. Last, a somewhat yet unexplored area, is the cross-linguistic representation one may try to analyze across languages. Translation is often full of nuance with many cultural considerations, so it may be difficult to map different language sets together to analyze the universality of concept representation. Further, even if they could be matched, different language norms are unlikely to contain exactly the same set of concepts, and the overlap between them may be low.

**Example Applications**

Feature production norms are valuable data that can be used across a variety of purposes. Psycholinguistic and cognitive psychology researchers are able to leverage the data for experimental control over stimuli presented in their study. For example, if you wanted to give participants related lists of concepts, the feature production norms could be used to calculate similarity between words to find the most related items. Semantic priming research investigates how similarity between pairs of words facilitates word recognition when the words are related versus unrelated (Meyer & Schvaneveldt, 1971). Feature lists represent an excellent source of data to create word pairs for both the related (i.e., share lots of features) and unrelated (i.e., share no features) conditions. Feature production norms also represent a rich source of data for secondary hypothesis that relate the feature values to other linguistic concepts. For example, Muraki et al. (2020) have investigated the relationship of these norms to semantic richness and theories of embodied cognition, which argue that knowledge representation is embedded with information about the physical body and movement within our environment (Barsalou, 1999). Therefore, when producing features for a concept like *run*, we use information about physically running to help determine those features, which in turn influences other behavior related to word recognition.

Figure 1 presents an early representation of how feature production norms were crucial to semantic network theory. The nodes (i.e., circles) in these models represent a concept or feature within the network, while the edges (i.e., lines) would represent the strength of connection between those features (Cree & Armstrong, 2012). The model, adapted from Collins and Loftus (1975), suggests that activation of any part of the network (e.g., through reading, hearing, thinking, etc.) would then spread outwards to other nodes in the network. For example, if one were to read the word *color*, it would then spread to *red,* and then *danger, blood,* and *primary*, and so on. Spreading activation theory, proposed in the 1970s (Collins & Loftus, 1975), heavily impacted the state and ideas behind computational modelling today, especially as semantic network theory and spreading activation inspired connectionism theories alternatives that large language models, such as BERT or ChatGPT, use as their framework today (Hinton & Anderson, 2014).



*Figure 1*. Feature relations from Buchanan et al. (2019) adapted from the proposed model by Collins and Loftus (1975). The original feature *red* is represented in the center and four one-hop features (*primary, danger, blood, color*) are shown on each side of *red.* One-hop networks represent the first features listed with a concept. The two-hop network is also presented, meaning, several most common features for the one-hop features and their interconnections (*blue, body, harm, yellow*, for example).

**Conclusion**

The feature production task is deceptively simple – what makes a *thing* that *thing*? The feature production norms created from this task represent an invaluable set of data for designing experiments, asking new research questions, and designing and testing models to represent semantic knowledge. The data can be contrasted with the free association norms that appear to capture experiential knowledge in addition to feature represented. Both types of data come with processing and analysis hurdles that face nearly all natural language processing tasks today where each choice should be considered carefully for potential impact on final results. The interest in building large language models that “think like humans” makes research on feature production especially interesting, as it provides a comparison for human knowledge representation. If you ask ChatGPT what makes a *cat* a cat, it suggests the following main points: Felidae Family (which includes the word *animal*, *domestic*), Physical Characteristics (*tail*), Whiskers, Carnivorous Diet, Behavioral Traits such as pouncing, Grooming, Communication (*meow*), and Domestication (*domestic*). Unlike humans, ChatGPT does not mention that cats have *fur*, are commonly *small,* and *pets*, and have *four legs.* These features are generally the most commonly listed features, which suggests that the feature production norms and research like it are still fruitful pieces of data to understand human knowledge.

**References**

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Buchanan, E. M., Deyne, S. D., & Montefinese, M. (2020). A practical primer on processing semantic property norm data. *Cognitive Processing*, *21*(4). https://doi.org/10.1007/s10339-019-00939-6

Buchanan, E. M., Valentine, K. D., & Maxwell, N. P. (2019). English semantic feature production norms: An extended database of 4436 concepts. *Behavior Research Methods*, *51*(4), 1849–1863. https://doi.org/10.3758/s13428-019-01243-z

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1

Cree, G. S., & Armstrong, B. C. (2012). Computational Models of Semantic Memory. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (1st ed., pp. 259–282). Cambridge University Press. https://doi.org/10.1017/CBO9781139029377.014

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research Methods*, *51*(3), 987–1006. https://doi.org/10.3758/s13428-018-1115-7

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–498. https://doi.org/10.3758/s13428-012-0260-7

Devereux, B. J., Tyler, L. K., Geertzen, J., & Randall, B. (2014). The Centre for Speech, Language and the Brain (CSLB) concept property norms. *Behavior Research Methods*, *46*(4), 1119–1127. https://doi.org/10.3758/s13428-013-0420-4

Hinton, G. E., & Anderson, J. A. (Eds.). (2014). *Parallel Models of Associative Memory* (0 ed.). Psychology Press. https://doi.org/10.4324/9781315807997

Kremer, G., & Baroni, M. (2011). A set of semantic norms for German and Italian. *Behavior Research Methods*, *43*(1), 97–109. https://doi.org/10.3758/s13428-010-0028-x

McRae, K., Cree, G. S., Seidenberg, M. S., & Mcnorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/BF03192726

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). Semantic memory: A feature-based analysis and new norms for Italian. *Behavior Research Methods*, *45*(2), 440–461. https://doi.org/10.3758/s13428-012-0263-4

Muraki, E. J., Sidhu, D. M., & Pexman, P. M. (2020). Mapping semantic space: Property norms and semantic richness. *Cognitive Processing*, *21*(4), 637–649. https://doi.org/10.1007/s10339-019-00933-y

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*(3), 506–515. https://doi.org/10.3758/BF03195597

Stein, L. M., & Gomes, C. F. D. A. (2009). Normas brasileiras para listas de palavras associadas: Associação semântica, concretude, frequência e emocionalidade. *Psicologia: Teoria e Pesquisa*, *25*(4), 537–546. https://doi.org/10.1590/S0102-37722009000400009

Taylor, K. I., Moss, H. E., & Tyler, L. K. (2007). The conceptual structure account: A cognitive model of semantic memory and its neural instantiation. In J. Hart & M. A. Kraut (Eds.), *Neural Basis of Semantic Memory* (1st ed., pp. 265–301). Cambridge University Press. https://doi.org/10.1017/CBO9780511544965.012

Tulving, E. (1972). Episodic and Semantic Memory. In E. Tulving, W. Donaldson, & G. H. Bower (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*(1), 183–190. https://doi.org/10.3758/BRM.40.1.183

Vivas, J., Vivas, L., Comesaña, A., Coni, A. G., & Vorano, A. (2017). Spanish semantic feature production norms for 400 concrete concepts. *Behavior Research Methods*. https://doi.org/10.3758/s13428-016-0777-2

**Relevant Websites (optional)**

* [www.wordnorms.com](http://www.wordnorms.com)
* <https://smallworldofwords.org>