

Semantic Networks and Prediction in Language Processing

Fanny Meunier¹, Raphaël Fargier¹, Vincent Voss²

¹ Université Côte d'Azur, CNRS, BCL, France

² Lexiflaire.fr, France

Contemporary theories of language processing converge on the view that comprehension is fundamentally predictive [1]. Rather than passively decoding input, the system continuously generates probabilistic hypotheses about upcoming information. Incoming input is evaluated against these expectations in real time, enabling efficient incremental interpretation and rapid adjustment when predictions are violated. Within this framework, meaning is not retrieved from a static lexicon but dynamically constructed through context-sensitive inference.

Semantic knowledge is typically modeled as a network in which word meanings are organized through associative links [2]. A standard method for probing such networks is the verbal association task, in which participants produce the first word that comes to mind in response to a stimulus. Large-scale association norms have been widely used to approximate the structure of the mental lexicon [3]. However, these paradigms largely abstract away from the interactive and anticipatory nature of language use, raising a critical question: do verbal association norms adequately capture the predictive mechanisms underlying real-time comprehension?

To address this issue, we compared semantic networks derived from traditional French verbal association norms with those emerging from an interactive, prediction-driven task. We analyzed data from a large-scale online word-guessing game (3,000 stimulus words; >10 million responses) involving two human participants: a proposer and a guesser [4]. The proposer provides up to four cue words within one minute to guide the guesser toward a target word. The dataset includes thousands of rounds across several thousand targets.

We compared two indicators to the ten associates in the normative database: (i) the first cue produced by proposers, and (ii) those that led directly to a correct guess. Crucially, when the guesser identifies the target from a single cue, this cue can be considered a highly efficient semantic predictor. Such trials resemble a verbal association task, as picked up by (i), with the key difference that the association must successfully support another individual's prediction under communicative constraints. Successful single-cue trials therefore index shared semantic knowledge mobilized in an explicitly predictive and interactive context (ii).

We analyzed first cues and compared them to the largest available French verbal association dataset (1,100 stimulus words; >89,000 responses) [5]. From their intersection, we selected 30 target words and compared normative associates with both the most frequent first cue overall and in successful single-cue trials.

At first glance, the results show only a partial overlap between the two datasets (see examples in 1): just 51% of norm responses appear among the game's top ten responses. However, when considering all game responses (not only the ten most frequent cues), 88% of the norm answers are found in the game. The remaining cases involve items excluded from the guessing task (e.g., non-dictionary words or morphologically related forms), and idiosyncratic responses such as *good-cherry*, which represented only 2.6%. An examination of cues from the top ten game responses indicates that 3 to 8 answers per target ($m = 4.8$) are absent from verbal norms, while sometimes ranking among the most frequent in the game ($m = 4$ th position). Interestingly, and contrary to expectations, shared associates between the norms and the game yield a proportion of direct correct guesses comparable to that of cues unique to the game (21% vs. 20%, respectively). This challenges the assumption that the verbal norm constitutes the best proxy for semantic networks, as this would predict a higher proportion of correct guesses than for cues unique to the game.

Overall, the findings suggest that existing French verbal association norms, beyond the first or second associate, lack sufficient lexical coverage, participant numbers, and trial density required to sustain experimental paradigms aimed at characterizing fine-grained semantic processing and its underlying representational structure. More fundamentally, they underscore the importance of incorporating interactive, prediction-based paradigms into models of semantic organization.

References

- [1] Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, 224, 105050. <https://doi.org/10.1016/j.cognition.2022.105050>
- [2] Deyne, S. D., Kenett, Y., Anaki, D., Faust, M., & Navarro, D. (2017). *Large-scale Network Representations of Semantics in the Mental Lexicon*. Psychology Press. <https://doi.org/10.4324/9781315413570-18>
- [3] Deese, J. (1966). *The structure of associations in language and thought*. Johns Hopkins Press.
- [4] Voss, V. (2015). Lexiflaire.fr.
- [5] Lakhzoum, D., Izaute, M., & Ferrand, L. (2023). Word-association norms for 1,100 French words with varying levels of concreteness. *Quarterly Journal of Experimental Psychology Series a Human Experimental Psychology*, 76(12), 2794–2803. <https://doi.org/10.1177/17470218231154454>

Targets & Translations	Associa- tive norms	Translations	Number of occur- rences	Cues - Guessing task	Translations	Number of occurrences	Number of good guesses af- ter the cue
pantalon trousers	jean	jeans	17	jean	jeans	1139	781
	short	shorts	11	short	shorts	782	327
	vêtement	clothing	9	vêtement	clothing	246	31
	jambe	leg	7	habit	garment	209	14
	ceinture	belt	3	jupe	skirt	141	11
	habit	garment	3	bas	low	122	0
	homme	man	3	chemise	shirt	82	7
	jupe	skirt	3	jambe	leg	68	1
	chemise	shirt	2	froc	pants	51	29
	découdre	unpick	2	jogging	sweatpants	50	6
marteau hammer	clou	nail	17	clou	nail	1049	513
	enclume	anvil	13	outil	tool	979	545
	piqueur	jackhammer	11	enclume	anvil	313	88
	outil	tool	6	piqueur	Jackhammer	165	60
	bricolage	DIY	4	tournevis	Screwdriver	165	53
	faucille	sickle	4	thor	Thor	159	65
	Thor	Thor	3	faucille	Sickle	156	38
	pioche	pickaxe	2	taper	Hit	85	1
	rouge	red	2	bricolage	DIY	53	10
	bruire	rumble	1	masse	sledgeham- mer	49	3
cheval horse	animal	animal	7	jument	Mare	605	473
	selle	saddle	5	poney	pony	594	505
	Troie	Troy	4	animal	animal	501	9
	âne	donkey	3	équitation	horse riding	192	163
	monter	ride	3	cavalier	rider	134	67
	ami	friend	2	galoper	gallop	111	71
	cavalier	rider	2	monture	mount	84	54
	chevalier	knight	2	canasson	nag	80	47
	course	race	2	poulain	foal	80	51
	étalon	stallion	2	horse	horse	71	61
chapeau hat	melon	bowler hat	20	casquette	hat cap	1149	687
	casquette	cap	15	tête	head	311	12
	soleil	sun	7	bonnet	beanie	257	71
	bas	low	6	melon	bowler hat	121	4
	pointu	pointy	6	sombrero	sombrero	108	66
	tête	head	6	bêret	beret	69	31
	chapi	-	3	coiffe	headdress	59	12
	haut-de- forme	top hat	3	accés- soire	accessory	53	0
	paille	straw	3	cowboy	cowboy	31	0
	Harry Potter	Harry Potter	2	bob	bucket hat	30	12

Figure 1: Examples of two words for which the responses in the two datasets are the most similar – seven shared responses (highlighted) –, and two words for which only two and three responses are shared.