Sapir's Thought-Grooves and Whorf's Tensors: Reconciling Transformer Architectures with Cultural Anthropology

Michael Castelle
Centre for Interdisciplinary Methodologies
University of Warwick
M.Castelle.1@warwick.ac.uk

Abstract

Contemporary deep learning technologies, from Transformer-based machine translation and multilingual speech recognition architectures to the CLIP model underpinning diffusion-based generative art, are increasingly viewed as being entangled with culture; however, computer science remains significantly detached from anthropology, the academic discipline with the most thorough conceptions of cultural action and, especially, the relationship of language and culture. How can these fields be reconciled? In this short paper I will suggest that the North American anthropologists E. Sapir and B.L. Whorf had a "high-dimensional" view of individual language interpretation and production which is far more relevant to Transformer-like architectures than has been previously appreciated (and certainly closer than that of the syntax-centric generative linguistics which largely supplanted early 20th-century anthropological paradigms of language). These intriguing parallels indicate that a far deeper synthesis of (already-multimodal) linguistic/cultural anthropology and contemporary AI research is possible, allowing for a) the de-centering of currently-influential but problematic cognitivist ideologies of language and thought and b) a more fine-grained and empirical critique of the social limitations of NLP architectures.

1 Introduction: The Semiotic Conception of Culture and its Relationship to Connectionist AI

"Culture" is not a word that ever gets defined within a typical undergraduate or graduate computer science curriculum; to some extent, it remains a fuzzy subject even in mainstream sociology, which eschewed broader theoretical conclusions from mid-20th-century studies of interpersonal interaction (Goffman (1959), Garfinkel (1967)) and for which the "sociology of culture" since Bourdieu (1984) has been dominated by quantitative studies of class and taste. Instead, the field in which the term "culture" is of most profound importance is anthropology, beginning with the late 19th-century cultural evolutionist Edward Tylor, who believed that societies passed through defined "stages" from "savagery" to civilization. However, a turn-of-the-century perspective in American anthropology, developed by the German-born American anthropologist Franz Boas, explicitly opposed the then-popular normative/ethnocentric views of culture (and/or *Kultur*) and steadfastly refused to privilege any one social grouping as intrinsically superior over another; it is with Boas that the term 'culture' began to be used in its now-common plural form, 'cultures' (Boas, 1896). For Boas' students, including Edward Sapir and Ruth Benedict, language was not only a prerequisite to culture, but an understanding of the diversity and lack of inherent superiority of specific human languages could be mapped onto new understandings of culture as well (Sapir, 1937; Benedict, 1934).

Preprint. Under review.

By the 1970s, influential anthropologists like Clifford Geertz were dedicating much of their career for a more ethnographically meaningful definition of culture, to a concept which he claimed was "essentially a semiotic one" (Geertz, 1973) but which was conceptually limited to a often-vague hermeneutical tradition (Parmentier, 1997). Instead, a more fully semiotic anthropology developed in the 1980s and 1990s which took on a distinctively processual view of discursive interaction, inspired not just by Boas and Sapir but also by C.S. Peirce, which centered neither syntax nor semantics but pragmatics and so-called metapragmatics (Silverstein, 1976). This holistic perspective of *language as social action* (Agha, 2006) is one which incorporates many reflexive language-based phenomena often ignored by linguists (e.g. citation, deixis, gesture, ritual) and thus brings forth a profound view of language which — unlike the views typically found in generative linguistics or natural language processing (NLP) — is closest to that of the modern conception of culture(s).

If, then, culture is intrinsically related to a reflexive (and recursive) semiotic discursivity, we can understand the developing interest in the relevance of culture with respect to the material techniques and technologies of contemporary AI research or *deep learning*, which is in part characterized by its distinctive layers of linear transformations and nonlinear rectifications with which it transforms its inputs. Linear transformations, when viewed geometrically, amount to simple rotations/reflections/scalings which preserve collinearity of high-dimensional vectors, and are thus a kind of *iconic* transformation in the sense of Peirce (1931); and the multiple, one-way layers of distortional rectifications in any forward pass of a neural network — which are not invertible — can be seen as an *indexical* flow (Weatherby and Justie, 2022). This connectionist revival differs strongly from previous "good old fashioned" as well as later statistical approaches to AI and NLP, which, appropriately, are referred to as *symbolic*, and thus dominated by *arbitrary* relations (in the Saussurean sense) between signifier and signified, as in the (in)famous Cyc project (Lenat and Guha, 1989).

For a semiotically-minded anthropologist, then, it is thoroughly unsurprising that "symbolic AI" was found to suffer from a "grounding problem" (Harnad, 1990). Nor is it surprising that the first late-1980s applications of multilayer neural networks to language data was met with reactionary dismissal by those with a strong investment in the assumptions of a dehumanizing algorithmic cognitivism (Pinker and Prince, 1988). Many aspects of contemporary neural networks, however, remain "ungrounded" in various key ways: beyond the obvious lack of embodiment (Haraway, 1988), the inert nature of most pretrained Transformer models means that they cannot accommodate the flux of experience (Bergson, 1911; French, 1999), and the inability to incorporate more than their prespecified maximum of input tokens means that access to interactional context is limited for generative models and chatbots (Thoppilan et al., 2022). Nevertheless, it is the case that at its technical core, the deep learning "revolution" of large-scale connectionism is, from an anthropological perspective, a kind of semiotic "revolution" which renders simplistic iconic-indexical transformations practical and effective for a variety of admittedly problematic and highly decontextualized tasks (Niven and Kao, 2019). Despite the best efforts of researchers, however, present-day deep learning models do not yet represent a *discursive* revolution which can account for or represent the many-layered processes of social action which can be empirically observed in human societies.

2 Linguistic Relativity as a Geometrical Transformation: Towards the Limitations of High-Dimensional Cultural Consilience

Boas' previously-mentioned form of cultural "relativity" would, in the case of language, be used by his protégé Sapir to argue against evolutionist views regarding then so-called "primitive" languages, arguing that "[w]hen it comes to linguistic form, Plato walks with the Macedonian swine-herd, Confucius with the head-hunting savage of Assam" (Sapir, 1921, p. 234). But Sapir would take linguistic relativity even further beyond this essential point. For him, to be able to interpret a sentence in spoken language — or to be able to express a thought in spoken language — required a mental process which was inevitably inflected by the obligatory and non-obligatory elements of that language's grammar. He wrote:

> "From the point of view of language, thought may be defined as the *highest latent or potential content of speech*, the content that is obtained *by interpreting each of the elements in the flow of language as possessed of its very fullest conceptual value*... [Language] humbly works up to the thought that is latent in, that may eventually be read into, its classifications and its forms; it is not, as is generally but

> naïvely assumed, the final label put upon the finished thought." (Sapir, 1921, p. 14) [emphasis added]

In this passage and others, the view of spoken language and thought that Sapir develops is one in which the act of creating speech 'forms' is seen as a collectively defined and largely unconscious art that unfolds in phonological time, a technique for taking listeners and speakers to-and-fro the latent 'content' of any given utterance. It differs significantly from the late-20th-century pronouncements of Pinker and Fodor, who believed that thinking explicitly occurred in a kind of combinatorial syntax-like 'mentalese' (a so-called "language of thought"). In addition, it implies that speakers of different languages might neccessarily follow different 'paths' between thoughts and speech, and by 1924, Sapir had drawn connections between this burgeoning linguistic "relativity" and the then-broadly-popular relativity of Einstein:

> "The world of linguistic forms, held within the framework of a given language, is *a complete system of reference*, very much as a number system is a complete system of quantitative reference *or as a set of geometrical axes of coordinates is a complete system of reference to all points of a given space*. The mathematical analogy is by no means as fanciful as it appears to be. *To pass from one language to another is psychologically parallel to passing from one geometrical system of reference to another...* the formal method of approach to the expressed item of experience, as to the given point of space, is so different that *the resulting feeling of orientation can be the same neither in the two languages nor in the two frames of reference.*" (Sapir, 1924, p. 153) [emphasis added]

Given this, Sapir's century-old view of "passing from one language to another" (e.g. as one would need to do in the act of translation) can be read as an admittedly-simplified, "single-layer" perspective of what actually happens in 21st-century neural machine translation (Johnson et al., 2017), as informed by the then-contemporaneous popularity of Einstein's special relativity. For Sapir, a multilingual speaker's mind moves along different well-worn "thought-grooves", as he called them, when converting an utterance into thought (or from thought to speech) in different languages.

One of Sapir's most dedicated students at Yale was a commuting Hartford insurance inspector with a MIT chemical engineering degree named Benjamin Lee Whorf, who would become an expert in the Maya writing system and the Uto-Aztecan family of languages. Whorf was not, despite his reputation, at all committed to any kind of conceptual incommensurability on the part of speakers of different languages; instead, he believed that as language users learn more ways to construct and express thoughts (perhaps through the learning of multiple languages, or through scientific study), the proverbial swineherd and headhunter might converse and come to agreement on abstruse philosophical topics, something the evolutionists had thought impossible.

In his final essay, "Language, Mind, and Reality", Whorf explicitly brings linguistic relativity to the mundane level of technical dialects, noting that "every language and every well-knit technical sublanguage incorporates certain points of view and certain patterned resistances to widely divergent points of view" — as we have indeed recently seen with the cultures of generative linguistics and connectionist NLP (Pater, 2019). For Whorf, such artificial resistances segregate the sciences, and also make impossible what he calls "the next great step in development". For his audience, the lay readers of *The Theosophist*, he proposes a radical vision — deliriously spiritual in its time, but today a surprisingly recognizable vision of unthinkably-vast multilayered and patterned relations (cf. Kovaleva et al. (2019)) in which, contemporary AI researchers hope, mathematics, art, and music will one day fall into a high-dimensional processual unity with the material tokens of linguistic form:

> "A noumenal world — a world of hyperspace, of higher dimensions — awaits discovery by all the sciences, which it will unite and unify, awaits discovery under its first aspect of a realm of *patterned relations*, inconceivably manifold and yet bearing a recognizable affinity to the rich and systematic organization of *language*, including *au fond* mathematics and music, which are ultimately of the same kindred as language... what I have called patterns are basic in a really cosmic sense, and that patterns form wholes, akin to the *Gestalten* of psychology which are embraced in larger wholes in continual progression. Thus the cosmic picture has a serial or hierarchical character, that of a progression of planes or levels." (Whorf, 1942, p. 248) [emphasis in original]

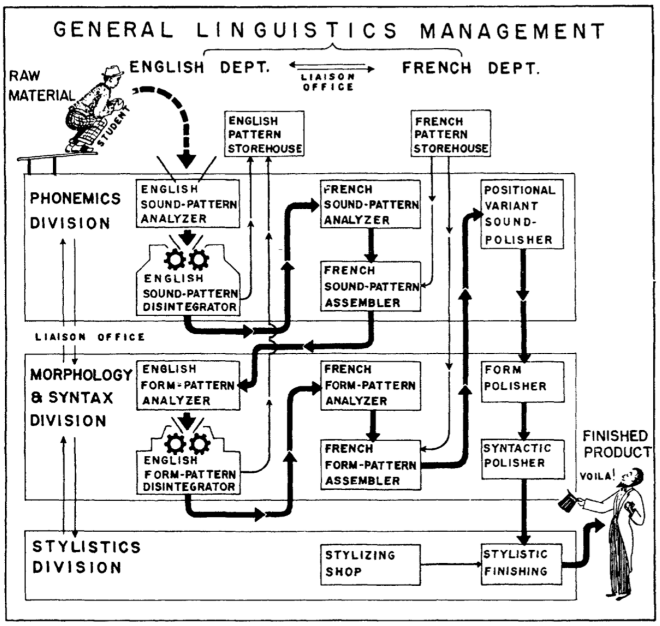

Figure 14. Flow sheet of improved process for learning French without tears. Guaranteed: no bottlenecks in production.

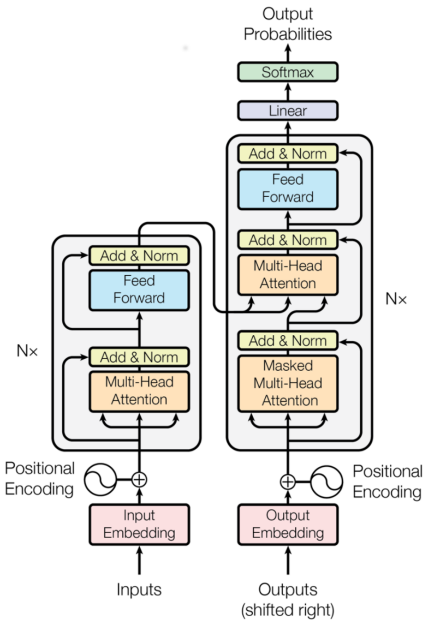

Figure 1: The Transformer - model architecture.

Figure 1: Left: Whorf describes the ideal process for learning how to translate English into French (Whorf, 1940). Right: Google describes the ideal process for learning how to translate English into French (Vaswani et al., 2017).

References

Asif Agha. 2006. *Language and Social Relations*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511618284

Ruth Benedict. 1934. *Patterns of culture*. Houghton Mifflin company, Boston, New York. OCLC: 608356.

Henri Bergson. 1911. *Creative Evolution*. London: Palgrave-Macmillan.

Franz Boas. 1896. The Limitations of the Comparative Method of Anthropology. *Science* 4, 103 (Dec. 1896), 901–908. https://doi.org/10.1126/science.4.103.901 Publisher: American Association for the Advancement of Science.

Pierre Bourdieu. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Harvard University Press.

Robert M. French. 1999. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences* 3, 4 (April 1999), 128–135. https://doi.org/10.1016/S1364-6613(99)01294-2

Harold Garfinkel. 1967. *Studies in ethnomethodology*. Prentice-Hall, Englewood Cliffs, N.J.

Clifford Geertz. 1973. *The interpretation of cultures: selected essays*. Basic Books, New York. OCLC: 737285.

Erving Goffman. 1959. *The presentation of self in everyday life* (anchor books edition ed.). Doubleday & Company, Garden City, N.Y. OCLC: 256298.

Donna Haraway. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14, 3 (1988), 575–599. https://doi.org/10.2307/3178066 Publisher: Feminist Studies, Inc..

Stevan Harnad. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42, 1 (June 1990), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the Association for Computational Linguistics* 5 (2017), 339–351. https://doi.org/10.1162/tacl_a_00065 Place: Cambridge, MA Publisher: MIT Press.

Olga Kovaleva, Alexey Romanov, Anna Rogers, and Anna Rumshisky. 2019. Revealing the Dark Secrets of BERT. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 4365–4374. https://doi.org/10.18653/v1/D19-1445

Douglas B. Lenat and R. V. Guha. 1989. *Building Large Knowledge-based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley Publishing Company. Google-Books-ID: 55NQAAAAMAAJ.

Timothy Niven and Hung-Yu Kao. 2019. Probing Neural Network Comprehension of Natural Language Arguments. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 4658–4664. https://doi.org/10.18653/v1/P19-1459

Richard J. Parmentier. 1997. The Pragmatic Semiotics of Cultures. *Semiotica* 116, 1 (1997), 1–114.

Joe Pater. 2019. Generative linguistics and neural networks at 60: Foundation, friction, and fusion. *Language* 95, 1 (2019), e41–e74. Publisher: Linguistic Society of America.

Charles S. Peirce. 1931. *Collected Papers of Charles Sanders Peirce*. Vol. 2.

Steven Pinker and Alan Prince. 1988. On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition. *Cognition* 28, 1-2 (1988), 73–193. https://doi.org/10.1016/0010-0277(88)90032-7

Edward Sapir. 1921. *Language: an introduction to the study of speech*. Harcourt, Brace, and Company, New York. OCLC: 853469.

Edward Sapir. 1924. The Grammarian and His Language. In *The selected writings of Edward Sapir in language, culture, and personality*, D.G. Mandelbaum (Ed.). University of California Press, Berkeley, 150–159.

Edward Sapir. 1937. Language. In *Encyclopaedia of the social sciences*, Edwin R. A. Seligman and Alvin Saunders Johnson (Eds.). Vol. 9. The Macmillan company, New York, 155–169. OCLC: 168443.

Michael Silverstein. 1976. Shifters, linguistic categories, and cultural description. In *Meaning in Anthropology*, Keith H. Basso and Henry A. Selby (Eds.). University of New Mexico Press, Albuquerque, 11–55.

Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, YaGuang Li, Hongrae Lee, Huaixiu Steven Zheng, Amin Ghafouri, Marcelo Menegali, Yanping Huang, Maxim Krikun, Dmitry Lepikhin, James Qin, Dehao Chen, Yuanzhong Xu, Zhifeng Chen, Adam Roberts, Maarten Bosma, Vincent Zhao, Yanqi Zhou, Chung-Ching Chang, Igor Krivokon, Will Rusch, Marc Pickett, Pranesh Srinivasan, Laichee Man, Kathleen Meier-Hellstern, Meredith Ringel Morris, Tulsee Doshi, Renelito Delos Santos, Toju Duke, Johnny Soraker, Ben Zevenbergen, Vinodkumar Prabhakaran, Mark Diaz, Ben Hutchinson, Kristen Olson, Alejandra Molina, Erin Hoffman-John, Josh Lee, Lora Aroyo, Ravi Rajakumar, Alena Butryna, Matthew Lamm, Viktoriya Kuzmina, Joe Fenton, Aaron Cohen, Rachel Bernstein, Ray Kurzweil, Blaise Aguera-Arcas, Claire Cui, Marian Croak, Ed Chi, and Quoc Le. 2022. LaMDA: Language Models for Dialog Applications. https://doi.org/10.48550/arXiv.2201.08239 arXiv:2201.08239 [cs].

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. (June 2017). https://arxiv.org/abs/1706.03762v5

Leif Weatherby and Brian Justie. 2022. Indexical AI. *Critical Inquiry* 48, 2 (Jan. 2022), 381–415. https://doi.org/10.1086/717312 Publisher: The University of Chicago Press.

Benjamin Lee Whorf. 1940. Linguistics as an Exact Science: New Ways of Thinking, Hence of Talking, about Facts Vastly Alter the World of Science, Emphasizing the Need for Investigation of Language. *MIT Technology Review* (Dec. 1940), 61–63, 80–83.

Benjamin Lee Whorf. 1956 [1942]. Language, Mind, and Reality. In *Language, Thought, and Reality*. MIT Press, 246–270.