

Ethnography as Language Model: Speaking Beyond the Text

Michael Fischer and Ben Kluga, Human Relations Area Files at Yale University

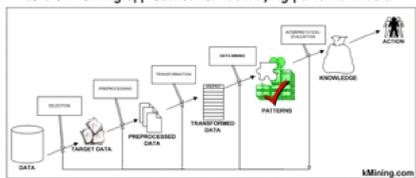


Ethnography

- Goals of ethnography - research record, results, communications
- Commonalities of ethnographic form - based on fieldwork, contemporary themes identified and documented, long term themes emerge
- Why use text vector representations of ethnography?
- How meaning is established in Word Vector Models and Language Models
- Applications - comparison, coding, theory building

DETECTING ORDER

- Data mining - knowledge from information
- Collecting Data
- Transforming Data
- Lots of working approaches for identifying patterns in data



Identifying significance

- ML/NLP evaluate the importance of a word to other words in a collection or corpus.
- Importance increases proportionally to the frequency of word clusters that appear in the ethnographies.
- NLP and ML tools identify patterns that emerge from relations between word clusters.



Goldilocks words - not too common or too rare

Word2Vec: method based on neural networks

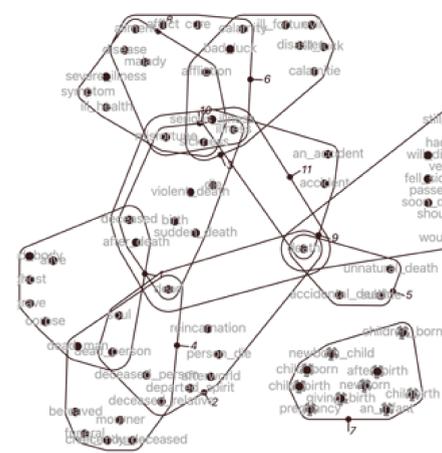
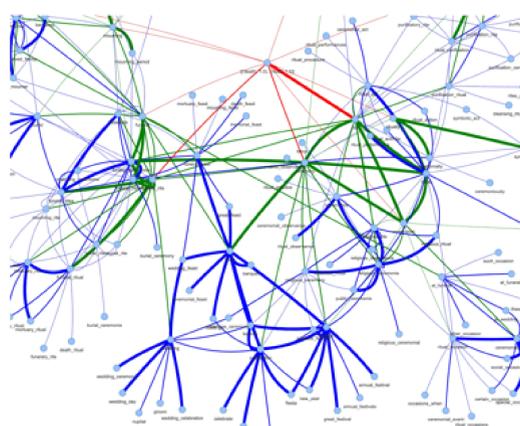
Culture	Intelligence	Myth
Word Similarity	Word Similarity	Word Similarity
cultural_tradition 0.86	intellect 0.77	mythology 0.89
cultural_pattern 0.84	capability 0.75	creation_myth 0.85
cultural 0.82	shrewdness 0.74	origin_myth 0.84
traditional_culture 0.82	aptitude 0.72	legend 0.84
other_culture 0.82	talent 0.72	tale 0.83
native_culture 0.81	intelligent 0.72	mythological 0.82
world_view 0.81	cleverness 0.72	origin_myths 0.80
subculture 0.80	superior_intelligence 0.71	mythical 0.79
cultural_system 0.79	intellectual_ability 0.71	mythic 0.77
western_culture 0.78	wisdom 0.70	these_story 0.77

Word2Vec: Magic and Power: Good vs Bad

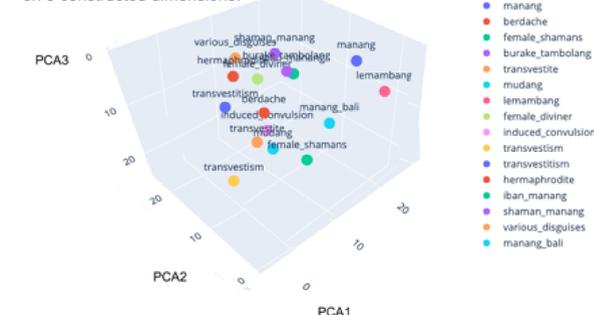
Word Similarity	Word Similarity	Word Similarity	Word Similarity
magical 0.72	black_magic 0.80	ability 0.75	powerless 0.71
charm 0.69	sovereignty 0.75	prestige 0.74	evil 0.69
magical_power 0.69	evil 0.74	wealth 0.72	fear 0.68
object 0.66	witchcraft 0.72	capacity 0.72	danger 0.64
magic_power 0.64	sovereign 0.70	political_power 0.71	evil_power 0.64
magical_charm 0.64	witch 0.70	control_ovr 0.70	cause 0.64
supernatural_power 0.64	evil_spirit 0.70	not_only 0.69	threat 0.62
medicine 0.64	spell 0.70	spiritual_power 0.69	machination 0.62
purpose 0.64	malignant 0.69	benefit 0.69	malvolence 0.62
requisite 0.63	magical 0.69	advantage 0.69	supernatural_power 0.61

Vocabulary Vector Functions

- Identifying related words within a domain, or between domains
- Exploring concepts through union and intersection of meanings
- Identifying terms for text search
- Locating indigenous terms within semantic space



3D Principal Components Plot: Relations between terms on 3 constructed dimensions.



Language Model Functions

- Summarizing ethnographic entries
- Identifying relational patterns within a set of entries
- Extracting structured information from entries
- Comparing the content of entries
- Identifying complex concepts
- Evaluating complex concepts

Example: Identifying elements for a model: Cayapa Marriage

- Gemini 2.5 (gemini.google.com - some free versions there as well)
- Initially we'll look at using Google's Gemini 2.5 LLM. This is a paid service, and one has to assume that any data uploaded to Gemini is not secure.
- Using HRAF I generated a copyright 'fair use' sample of text on marriage by searching for OCM identifiers relating to Marriage (800-809) for the Cayapa (OWC: SD06) and saving the results as a text file.
- The resulting text file can be accessed at [sd06_all_marriage](#), and you are free to use it for your own purposes.
- The following slides will cover dialogs with LLMs based on this content.



QR Link to Presentation and Links to Examples

Gemini2.5 First contact:

the following text is about marriage among the Cayapa people. List what we can infer from the text, including deontic transitions and deontic themes. [text file submitted see [sd06_all_marriage](#)]

Gemini: Here's a breakdown of the inferences we can make from the provided text about Cayapa marriage, including deontic themes and transitions:

- "Kinship & Descent":
 - Bilateral descent: Kinship is traced through both the mother's and father's lines. (sd06-000-000182, sd06-004-001289)
 - Kinship-based: The main kin group is the kindred (relatives on both sides).
 - Shifting Kinship Terminology: There's evidence of a shift in kinship terminology between 1900 and 1960, possibly related to tribal fission and changes in residence. (sd06-004-001185)
 - Inequid Cousin Terminology: Cross-cousins are equated and parallel cousins are distinguished. (sd06-004-001253)
- "Residence":
 - Primarily bilocal: Couples tend to live near either the husband's or the wife's family. (sd06-000-000182, sd06-004-001289)
 - Some neolocality: Some couples establish a new, independent residence. (sd06-000-000182, sd06-004-001289)
- "Social Organization & Governance":
 - Self-governing: The Cayapa largely manage their own affairs.

Gemini Example: Outcomes in a nutshell

Fischer: the following text is about marriage among the Cayapa people. List what we can infer from the text, including deontic transitions and deontic themes. [text file submitted see [sd06_all_marriage](#)]

Fischer: elaborate on Marriage disputes: Conflicts arise over marriage arrangements, choice of partners, adherence to rules.

Fischer: elaborate on Daughter's Age: The mother may feel her daughter is too young. (sd06-004-004412)

Fischer: give some examples of Violation of norms to punishment:

Fischer: are there any candidate cultural models that have sufficient detail that would benefit from simulation?

Fischer: design a simulation using production rules, where each rule is of the form [Condition Outcome]. Condition can be a specific event, a state, or during the duration of an action. Outcome is an modal operator, Obligated, Permitted or the negation of these, followed by designation of an action or state. Devise other forms if needed. Do this for the simulation of: Marriage Management Process (with Reluctance & Negotiation)

Fischer: develop an interpreter written in prolog that can execute the production rules listed before.

Fischer: please finish the code

Fischer: show the code for "find_eligible_partner"

Fischer: Now redo the simulation in Python instead of Prolog.

Small LLM: Gemma3 (4b)

the following text is about marriage among the Cayapa people. List important specific inferences from the text, especially deontic transitions and deontic themes. [Text here ... see [sd06_all_marriage](#)]

Transcript for Gemma3 (19 pages)

Gemma3:

- "Overall Themes & Context:"
 - "Cayapa Tribe Culture:" The text describes the culture of the Cayapa tribe, a group of indigenous people likely in the region of Ecuador (referenced in some segments).
 - "Marriage Customs:" The core of the text revolves around their marriage customs, particularly the "boda" - a celebration involving feasts, dancing, and often, rituals relating to the deceased.
 - "Traditional Law & Social Control:" The Cayapa have strict rules regarding consanguineal marriage (marriage between relatives) and divorce. These laws are enforced by tribal officials.
 - "Social Dynamics:" The text highlights the complexities of social control, the role of women in the community, and the potential for deviation from established norms.
- "Key Observations & Summaries by Section:"
 - "Sections 1-4 (Focus on General Culture & Marriage):" These sections establish the basic cultural context. We learn that "boda" celebrations are common, encompassing both feasts and rituals related to the dead. There's a strong emphasis on family ties and the prohibition of marriage between relatives. The existence of divorce is acknowledged, but tightly regulated and often leading to exile.
 - "Sections 5-8 (The Woman's Case - A Specific Event):"

Gemma3 Example: Outcomes in a nutshell

Fischer: the following text is about marriage among the Cayapa people. List important specific inferences from the text, especially deontic transitions and deontic themes. [Text appears here ... see [sd06_all_marriage](#)]

Fischer: create a detailed list of deontic transitions in the text, and the agent roles relating to these transitions.

Fischer: analyze the reasoning behind the transitions.

Fischer: produce a prolog program that states the conclusions, conditions and rules above. The purpose will be to develop a simulation based on production rules.

Fischer: expand the program with more rules and facts.

Fischer: add some facts and rules relating to arrangement of marriages

Fischer: introduce variables to track the relationships between families and the state of the marriage proposals.

Discussion

1. The large and small models are comparable. Both will help you with rapidly achieving around 90% of your task.
2. On both you have to exercise critical judgement.
3. The larger one is a bit more forthcoming, while the small model is more terse, but both will do the same job if you direct them to.
4. The larger model may be slightly more accurate and write better simulations, though I don't usually find the difference compelling.
5. The small model will run on just about any recent laptop or desktop with 8 gig of ram at a speed adequate for assisting with scholarship.
6. The small model is private - your data will go no further than your own computer.
7. The smaller models will even run, if a bit slowly, on minimal hardware, such as a Raspberry Pi ...

Thanks!

Michael Fischer
Human Relations Area Files
Yale University
mike@hraf.net
<https://ethnolab.hraf.net>
Link to this presentation at:
<https://ethnolab.hraf.net>



This material is based upon work supported by the National Science Foundation under Grant Number 2024286. HDNS-I: Infrastructure for Knowledge Linkages from Ethnography of World Societies. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

This material is based upon work supported by the National Science Foundation under Grant Number 2024286. HDNS-I: Infrastructure for Knowledge Linkages from Ethnography of World Societies. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

