26º ENANCIB
Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação
A Amazônia na sociedade global em rede: assimetrias e potencialidades informacionais, tecnológicas e ambientais
Belém, Pará, 09 a 13 de novembro de 2026
Universidade Federal do Pará

XXVI ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO - XXVI ENANCIB

ISSN 2177-3688

GT 2 – Organização e Representação do Conhecimento

TESAURO E IA: O PAPEL DA ESTRUTURA SEMÂNTICA NA INTERPRETAÇÃO

THESAURI AND AI*: THE ROLE OF SEMANTIC STRUCTURE IN INTERPRETATION*

Benildes Coura Moreira dos Santos Maculan – Universidade Federal de Minas Gerais (UFMG)
Elisângela Cristina Aganette – Universidade Federal de Minas Gerais (UFMG)
Gercina Ângela de Lima – Universidade Federal de Minas Gerais (UFMG)
Frederico Mafra Pereira – Universidade Federal de Minas Gerais (UFMG)

Modalidade: Trabalho Completo

Resumo: A incorporação de modelos de linguagem baseados em Inteligência Artificial em atividades de busca, recuperação e representação da informação tem ampliado discussões sobre semântica, diversidade cultural e mediação conceitual em Sistemas de Organização do Conhecimento. Este estudo busca analisar, sob a perspectiva da Organização do Conhecimento, as diferenças nos processos de interpretação conceitual e representação temática produzidos por um modelo de linguagem quando opera com e sem acesso à estrutura semântica de um tesauro culturalmente situado, tomando como domínio empírico o patrimônio cultural amazônico representado no Yãkoana – Tesauro dos Bens Culturais do Amazonas. A pesquisa, de natureza qualitativa, exploratória e analítica, realizou um experimento com o GPT-5 em três condições: interpretação de termos sem acesso ao tesauro, interpretação mediada por sua estrutura semântica e representação temática de textos mediante seleção de descritores. A análise temática reflexiva identificou três dimensões analíticas: preservação conceitual, preservação do contexto sociocultural e mobilização das relações semânticas do vocabulário. Os resultados indicam que a disponibilização da estrutura semântica esteve associada a interpretações mais delimitadas e contextualizadas e forneceu referências terminológicas e relacionais para a representação temática. Evidenciaram também que a conformidade dos descritores com o vocabulário não assegura, por si só, sua pertinência temática em relação ao conteúdo dos documentos. Conclui-se que vocabulários controlados podem atuar como estruturas de mediação semântica em aplicações baseadas em modelos de linguagem, orientando os processos de interpretação e representação sem determinar integralmente as respostas produzidas.
Palavras-chave: organização do conhecimento; inteligência artificial; patrimônio cultural.

Abstract: The incorporation of Artificial Intelligence-based language models into information search, retrieval, and representation activities has broadened discussions on semantics, cultural diversity, and conceptual mediation in Knowledge Organization Systems. This study seeks to analyze, from a Knowledge Organization perspective, differences in the processes of conceptual interpretation and subject representation produced by a language model when operating with and without access to the semantic structure of a culturally situated thesaurus, taking as its empirical domain the Amazonian cultural heritage represented in the Yãkoana – Thesaurus of the Cultural Heritage of Amazonas. The qualitative, exploratory, and analytical study conducted an experiment with GPT-5 under three conditions: interpretation of terms without access to the thesaurus, interpretation mediated by its semantic structure, and subject representation of texts through descriptor selection. Reflexive thematic analysis identified three analytical dimensions: conceptual preservation, preservation of sociocultural context, and mobilization of the vocabulary's semantic relations. The results indicate that the availability of the semantic structure was associated with more clearly delimited and contextualized interpretations and provided terminological and relational references for subject representation. They also showed that the conformity of descriptors with the vocabulary does not, in itself, ensure their thematic relevance to the content of the documents. It is concluded that controlled vocabularies can function as structures of semantic mediation in language model-based applications, guiding interpretation and representation processes without fully determining the responses produced.
Keywords: knowledge organization; artificial intelligence; cultural heritage.

1 INTRODUÇÃO

A crescente incorporação de modelos de linguagem baseados em Inteligência Artificial (IA) em atividades de busca, recuperação e produção de informação tem ampliado discussões sobre representação semântica, vieses linguísticos e assimetrias culturais em sistemas automatizados de processamento da linguagem (Gallegos et al., 2024). No campo da Organização do Conhecimento (OC), essas discussões tornam-se particularmente relevantes diante do reconhecimento de que processos de representação e organização da informação dependem de estruturas conceituais, relações semânticas e contextos socioculturais específicos (Olson, 2002; Hjørland, 2008; Drabinski, 2009).

Em domínios culturalmente situados, como aqueles relacionados ao patrimônio cultural amazônico, a representação de conhecimentos locais e terminologias especializadas depende tanto da recorrência dos termos em grandes volumes de dados textuais quanto de estruturas semânticas capazes de preservar relações conceituais e contextos culturais associados aos domínios representados. Contudo, a diversidade cultural e linguística da Amazônia nem sempre se encontra adequadamente representada em sistemas de organização do conhecimento (SOC)

generalistas e em bases de dados utilizadas no treinamento de modelos de IA, favorecendo processos de generalização semântica, apagamento contextual e redução interpretativa.

Nesse contexto, tesauros especializados podem atuar como instrumentos de mediação semântica capazes de oferecer maior contextualização terminológica para sistemas baseados em IA. Entre esses instrumentos, destaca-se o Yãkoana – Tesauro dos Bens Culturais do Amazonas, desenvolvido para representar bens culturais materiais e imateriais, práticas sociais, saberes tradicionais e manifestações culturais amazônicas a partir de relações semânticas culturalmente situadas.

Apesar do crescimento das discussões sobre IA, vieses algorítmicos e diversidade linguística, ainda são limitadas, sob a perspectiva da OC, as investigações que comparam processos de interpretação e representação realizados por modelos de linguagem com e sem apoio de estruturas semânticas especializadas e culturalmente situadas. Assim, este estudo busca analisar, sob a perspectiva da Organização do Conhecimento, as diferenças nos processos de interpretação conceitual e representação temática produzidos por um modelo de linguagem quando opera com e sem acesso à estrutura semântica de um tesauro culturalmente situado, tomando como domínio empírico o patrimônio cultural amazônico representado no Yãkoana – Tesauro dos Bens Culturais do Amazonas.

2 REFERENCIAL TEÓRICO

O referencial teórico deste estudo articula contribuições estruturadas em três eixos complementares: (i) OC e estruturação semântica; (ii) garantia cultural, contexto e assimetrias na representação; e (iii) modelos de linguagem e representação semântica, com discussões acerca de vieses semânticos em IA e em SOCs.

A OC dedica-se à estruturação, representação e mediação do conhecimento em diferentes contextos informacionais. Para Dahlberg (1993), os conceitos constituem a unidade central da OC, uma vez que a representação depende da explicitação de unidades conceituais e de suas relações. Hjørland (2008), por sua vez, compreende a OC como uma prática relacionada a diferentes concepções de conhecimento, linguagem e organização social, envolvendo decisões semânticas e epistemológicas influenciadas por contextos socioculturais específicos. As duas perspectivas convergem quanto à centralidade da organização conceitual, mas diferem na ênfase: enquanto Dahlberg privilegia a estruturação dos conceitos e de suas relações, Hjørland evidencia as condições epistemológicas e sociais que participam dessa estruturação.

Para este estudo, essa distinção é relevante porque permite considerar simultaneamente a estrutura semântica do tesauro e sua vinculação a um domínio cultural específico. A representação da informação, portanto, não constitui uma atividade neutra, mas um processo de mediação intelectual e cultural.

A dimensão semântica da representação torna-se particularmente relevante em ambientes computacionais. Segundo Specia e Rino (2002), a representação semântica corresponde à formalização do significado por meio de estruturas conceituais e relações entre conceitos, não decorrendo apenas da presença isolada de termos. Nessa perspectiva, os SOCs atuam como mecanismos de mediação semântica, ao explicitar relações conceituais e contextualizar significados que orientam processos de descrição, associação e recuperação da informação. Em ambientes baseados em IA, tais estruturas podem complementar os padrões estatísticos mobilizados pelos modelos de linguagem.

A construção desses sistemas envolve também critérios de legitimação terminológica. Nesse contexto, a noção de garantia constitui um dos fundamentos epistemológicos da OC (Beghtol, 1986). O termo “garantia literária” foi formulado por Hulme (1911) e, posteriormente, o princípio de garantia foi ampliado por diferentes autores, dando origem a outras modalidades, como as garantias de usuário, organizacional, acadêmica e cultural (Barité, 2009, 2011, 2022). Entre elas, a garantia cultural assume especial relevância ao considerar que a representação do conhecimento deve respeitar as formas pelas quais diferentes comunidades interpretam, nomeiam e organizam a realidade. Em perspectiva semelhante, Beghtol (1995, 2002) defende que os sistemas de representação sejam capazes de acolher múltiplas perspectivas culturais e linguísticas, reduzindo processos de exclusão e apagamento semântico. Essas contribuições convergem ao questionar a suficiência de uma legitimação terminológica fundada apenas na recorrência documental: a representação de domínios culturalmente situados exige considerar também os contextos nos quais os conceitos adquirem significado. Nesse sentido, a garantia cultural não se opõe à estruturação semântica, mas acrescenta a ela uma exigência de adequação contextual, particularmente relevante para o domínio investigado neste estudo.

Essas discussões tornam-se particularmente relevantes diante da crescente utilização de modelos de linguagem de grande escala (Large Language Models – LLMs). Segundo Russell e Norvig (2022), a IA contemporânea passou a incorporar preocupações relacionadas à representação do conhecimento, linguagem natural, ética, confiabilidade e segurança. Entretanto, como observam Gallegos *et al*. (2024), tais modelos podem reproduzir vieses e

assimetrias presentes nos dados utilizados em seu treinamento, favorecendo terminologias mais recorrentes em detrimento de expressões locais ou culturalmente especializadas. Essa questão aproxima-se das críticas de Olson (2002) e Drabinski (2009), que destacam o caráter não neutro dos SOCs e sua relação com estruturas de poder e visibilidade informacional. Há, contudo, uma diferença importante entre os dois problemas: nos modelos de linguagem, as assimetrias podem decorrer da distribuição estatística e da composição dos dados de treinamento; nos SOCs, decorrem também de decisões explícitas de seleção, categorização e estabelecimento de relações conceituais. Assim, recorrer a um SOC para orientar um modelo de linguagem não elimina o problema da representação, mas desloca parte da mediação semântica para uma estrutura conceitual construída e passível de exame crítico.

No âmbito internacional, a UNESCO (2022) enfatiza a necessidade de promover diversidade cultural e pluralidade linguística no desenvolvimento de sistemas inteligentes. Em direção semelhante, Freitag *et al*. (2025) defendem que iniciativas voltadas à construção de uma IA brasileira devem contemplar a diversidade linguística e cultural do país. Em domínios culturalmente situados, como o patrimônio cultural amazônico, essas questões reforçam a importância de estruturas semânticas especializadas capazes de oferecer maior contextualização terminológica e apoiar processos automatizados de interpretação e representação do conhecimento.

3 METODOLOGIA

A pesquisa caracteriza-se como qualitativa, exploratória e analítica, orientada pela perspectiva da OC aplicada à representação de conhecimentos culturalmente situados. O corpus reúne termos extraídos do Yãkoana – Tesauro dos Bens Culturais do Amazonas (Trindade, 2025), vocabulário controlado voltado à representação de bens culturais materiais e imateriais, saberes tradicionais e manifestações culturais da região amazônica. A escolha do Yãkoana deve-se à sua aderência ao domínio investigado e à explicitação de categorias e relações semânticas vinculadas a um contexto sociocultural específico.

A seleção dos cinco termos foi intencional e orientada por critérios de diversidade conceitual e categorial. Buscou-se contemplar diferentes níveis de abstração e três eixos macrotemáticos do domínio: (a) dimensão imaterial e cognitiva, representada por Conjunto de Saberes e Língua Indígena; (b) dimensão expressiva e social, representada por Manifestação Cultural; e (c) dimensão territorial e material, representada por Encontro das Águas dos Rios

Negro e Solimões e Relógio Municipal. A composição permitiu contemplar conceitos abstratos e culturalmente dependentes, bem como termos com maior ancoragem material, territorial e geográfica. A seleção teve finalidade analítica, não se destinando à representatividade estatística do vocabulário. Os termos foram extraídos da versão do Tesauro Yãkoana publicada em 2025.

Para a Condição C, foram utilizados três textos curtos, com extensão entre 147 e 184 palavras, elaborados mediante adaptação de conteúdos publicados no portal Ateliê Amazônico (ATELIÊ AMAZÔNICO, [2026?]). A utilização de conteúdos provenientes de fonte externa ao tesauro buscou evitar circularidade entre a estrutura do vocabulário e os documentos submetidos à representação temática. Foram contemplados conteúdos relacionados a línguas indígenas, manifestações culturais e patrimônio natural, possibilitando observar a seleção de descritores diante de diferentes conteúdos documentais.

O experimento foi conduzido com o modelo de linguagem GPT-5, acessado por meio da plataforma ChatGPT, na versão disponibilizada pela OpenAI em maio de 2026, em três condições: (a) interpretação de termos sem fornecimento de definições ou relações semânticas extraídas do vocabulário controlado, embora o prompt explicitasse o contexto amazônico; (b) interpretação dos mesmos termos com acesso à estrutura semântica do tesauro; e (c) representação temática de textos mediante seleção de descritores do vocabulário. Na Condição B, foram fornecidas ao modelo as relações semânticas e, quando disponíveis, as notas de escopo constantes do Yãkoana. Na Condição C, o modelo recebeu o tesauro e foi instruído a selecionar exclusivamente descritores nele constantes.

Os prompts foram padronizados para cada condição e executados em sessões independentes, sem encadeamento de contexto entre as interações. Para cada termo ou texto, foi realizada uma única execução do prompt correspondente. Os prompts, os materiais fornecidos ao modelo e as respostas integrais encontram-se documentados nos anexos: o Anexo A reúne os textos e as respostas da Condição C; o Anexo B, as respostas da Condição A; e o Anexo C, as estruturas semânticas fornecidas e as respostas da Condição B.

O percurso metodológico compreendeu quatro etapas: (1) seleção e preparação do corpus, envolvendo a seleção dos cinco termos do Yãkoana e a preparação dos três textos utilizados na tarefa de representação temática; (2) realização das interações com o modelo nas três condições e organização das respostas por condições experimentais; (3) construção

iterativa das categorias analíticas, mediante leitura interpretativa e comparação das respostas; e (4) síntese interpretativa dos padrões identificados nas diferentes condições.

As respostas produzidas pelo modelo constituíram o corpus analítico, tendo como unidades de análise as interpretações conceituais e as representações temáticas produzidas nas diferentes condições. A análise temática reflexiva (Braun; Clarke, 2019; 2021) orientou a interpretação qualitativa das respostas geradas. A construção das categorias ocorreu de forma iterativa, em articulação com o referencial da OC. As categorias não foram estabelecidas a priori como uma grade fixa de codificação nem derivadas de procedimentos de frequência ou recorrência lexical. O processo envolveu a identificação, comparação e o refinamento progressivo de padrões relacionados à preservação e ao deslocamento conceitual, à contextualização sociocultural e à mobilização das relações semânticas do vocabulário. As dimensões resultantes desse processo são apresentadas na seção de resultados.

O estudo não se orientou por métricas quantitativas de desempenho, mas por uma abordagem interpretativa voltada à compreensão das relações entre estrutura semântica, contexto sociocultural e processos automatizados de interpretação e representação.

4 RESULTADOS E DISCUSSÃO

A primeira etapa concentrou-se na definição e preparação do corpus experimental a partir do Tesauro Yãkoana. Considerando o caráter qualitativo e interpretativo do estudo, foram selecionados cinco termos de forma intencional, buscando contemplar diversidade conceitual e categorial e diferentes níveis de abstração. A seleção abrangeu dimensões relacionadas à diversidade linguística, conhecimento tradicional, manifestações culturais, territorialidade e patrimonialização. O Quadro 1 apresenta os termos selecionados e suas respectivas justificativas analíticas.

Quadro 1 – Termos selecionados do tesauro Yãkoana e respectivas justificativas analíticas

Tipo	Termo	Justificativa analítica
Imaterial	LÍNGUA INDÍGENA	Diversidade linguística e representação culturalmente situada
Imaterial	CONJUNTO DE SABERES	Abstração conceitual e conhecimento tradicional
Imaterial	MANIFESTAÇÃO CULTURAL	Expressões e práticas culturais coletivas

Natural	ENCONTRO DAS ÁGUAS DOS RIOS NEGRO E SOLIMÕES	Territorialidade e contextualização cultural
Material	RELÓGIO MUNICIPAL	Patrimonialização e especificidade contextual

Fonte: elaboração própria (2026)

O conjunto selecionado reuniu, assim, diferentes configurações conceituais do Tesauro Yãkoana, abrangendo conceitos abstratos e culturalmente dependentes, manifestações socioculturais e bens de maior ancoragem territorial e material. Essa diversidade constituiu a base para a comparação das respostas produzidas nas diferentes condições experimentais.

A **segunda etapa** concentrou-se nas interações entre o modelo e os termos selecionados do Tesauro Yãkoana em três condições experimentais, definidas a partir de diferentes formas de acesso à estrutura semântica do vocabulário. As interações foram conduzidas por meio de *prompts* padronizados, executados em sessões independentes e iniciadas sem histórico conversacional prévio, de modo a reduzir possíveis efeitos de contexto acumulado e favorecer maior independência entre as diferentes condições experimentais. As respostas produzidas foram sistematizadas por cada condição experimental, constituindo o *corpus* analítico utilizado nas etapas posteriores de construção das categorias e síntese interpretativa.

Na **condição experimental A da segunda etapa**, os termos selecionados foram apresentados ao modelo sem fornecimento de definições, relações semânticas ou estrutura terminológica derivada do tesauro, embora o *prompt* explicitasse o contexto amazônico. O objetivo foi observar as associações conceituais e relações semânticas mobilizadas pelo modelo sem acesso à estrutura do vocabulário.

Para esta condição, utilizou-se o seguinte *prompt*: “Explique o que é [termo] no contexto amazônico e indique cinco termos semanticamente relacionados, justificando brevemente cada relação estabelecida.” O Quadro 2 apresenta as respostas das interações.

Quadro 2 – Síntese das respostas geradas pelo modelo para os termos analisados na condição A

Termo	Síntese da resposta do modelo
Língua indígena	Definido como sistema linguístico associado aos povos originários da Amazônia, relacionado à transmissão de conhecimentos, memória coletiva, identidade cultural e território. Os termos associados enfatizaram patrimônio

	cultural, tradição oral, identidade étnica e diversidade linguística.
Conjunto de saberes	Definido como um conjunto de conhecimentos, práticas e crenças transmitidos entre gerações e vinculados à experiência comunitária, ao território e à memória coletiva. Os termos associados enfatizaram tradição oral, sabedoria popular, prática artesanal, medicina tradicional e patrimônio cultural imaterial.
Manifestação cultural	Definido como expressão coletiva de identidades, práticas sociais, crenças e formas de vida. As associações estabelecidas enfatizaram celebrações religiosas, línguas indígenas, conjuntos de saberes, práticas artesanais e patrimônio cultural, sugerindo uma interpretação ampla do conceito vinculada à diversidade sociocultural amazônica.
Encontro das Águas dos Rios Negro e Solimões	Definido como fenômeno natural resultante da confluência dos rios Negro e Solimões. As relações estabelecidas enfatizaram elementos geográficos e hidrológicos, incluindo os rios envolvidos, o Rio Amazonas, Manaus e a noção de patrimônio natural.
Relógio Municipal	Definido como bem patrimonial urbano associado à memória coletiva, à identidade local e à paisagem cultural das cidades. As associações estabelecidas enfatizaram patrimônio material, praça pública, centro histórico, arquitetura urbana e memória coletiva, refletindo uma interpretação inferencial do conceito.

Fonte: elaboração própria (2026)

As respostas geradas constituíram a base empírica para a análise das dimensões de preservação conceitual, preservação do contexto sociocultural e mobilização das relações semânticas exposta na Etapa 4.

Na **condição experimental B da segunda etapa**, o modelo recebeu os termos acompanhados das relações semânticas e, quando disponíveis, das notas de escopo extraídas do Tesauro Yãkoana. O objetivo foi observar diferenças nos processos de interpretação e associação conceitual quando a estrutura semântica do vocabulário estava disponível ao modelo. O *prompt* utilizado foi estruturado da seguinte forma: "Considere a seguinte entrada do Tesauro Yãkoana. Com base exclusivamente na estrutura semântica apresentada (termo geral, termos específicos e termos relacionados, quando disponíveis), explique o conceito [termo] e descreva os vínculos semânticos que podem ser inferidos a partir dessa organização." O Quadro 3 sintetiza as respostas das interações.

Quadro 3 – Síntese das respostas geradas pelo modelo para os termos analisados na condição B

Termo	Síntese da resposta do modelo
Língua indígena	Interpretada como categoria de bem imaterial composta por diferentes línguas específicas. O modelo utilizou explicitamente as relações hierárquicas e associativas do tesauro para caracterizar o conceito e inferir vínculos com processos de reconhecimento, registro e preservação cultural.
Conjunto de saberes	Interpretado como categoria de bem imaterial associada a conhecimentos e práticas transmitidos socialmente. A explicação foi construída a partir dos termos específicos do tesauro, relacionados a sistemas agrícolas, cultivos tradicionais e saberes territorialmente situados na Amazônia, reforçando a dimensão cultural e comunitária do conceito.
Manifestação cultural	Interpretada como categoria de bem imaterial que reúne diferentes formas de expressão cultural. A explicação foi construída a partir dos termos específicos do tesauro, incluindo manifestações festivas, religiosas, artísticas, educativas e comunicacionais, o que conferiu maior contextualização e aderência ao patrimônio cultural amazônico.
Encontro das Águas dos Rios Negro e Solimões	Interpretado como patrimônio natural reconhecido por seu valor ambiental, paisagístico e cultural. A resposta foi estruturada a partir da nota de escopo e dos termos relacionados do tesauro, enfatizando os mecanismos institucionais de registro, tombamento e proteção patrimonial associados ao bem.
Relógio Municipal	Interpretado como bem patrimonial integrante do Centro Histórico de Manaus. A resposta utilizou a nota de escopo e as relações hierárquicas, partitivas e associativas do tesauro para contextualizar historicamente o bem e vinculá-lo aos processos institucionais de tombamento e proteção patrimonial.

Fonte: elaboração própria (2026)

As respostas produzidas constituíram a base empírica para a análise comparativa apresentada na Etapa 4.

Na **condição experimental C da segunda etapa**, utilizou-se a versão integral do Tesauro Yãkoana (Trindade, 2025), permitindo aproximar a tarefa de um contexto de representação temática orientada por vocabulário controlado, preservando o acesso do modelo à estrutura completa do tesauro.

Para essa condição, utilizou-se o seguinte *prompt*: “Com base exclusivamente no Tesauro Yãkoana fornecido em anexo, indique os cinco descritores mais representativos do texto apresentado. Os descritores devem constar no tesauro. Ordene-os por relevância temática e justifique brevemente cada escolha, mencionando, quando pertinente, as relações semânticas do vocabulário utilizadas na seleção.”

Foram utilizados três textos relacionados a diferentes conteúdos documentais: diversidade linguística indígena, manifestação cultural e patrimônio natural. O Quadro 4 sintetiza seus conteúdos e os descritores selecionados pelo modelo.

Quadro 4 – Conteúdo dos textos analisados na condição C e descritores selecionados pelo modelo

Texto	Conteúdo principal	Descritores selecionados
Texto 1	Diversidade linguística indígena amazônica, perda histórica de línguas, transmissão de conhecimentos tradicionais e políticas de reconhecimento e preservação linguística.	Língua Indígena; Política Estadual de Proteção das Línguas Indígenas do Estado do Amazonas; Registro de Bens Culturais de Natureza Imaterial; Yanomami; Nheengatu.
Texto 2	Maracatu como manifestação cultural afro-brasileira, articulando música, religiosidade, simbolismo, memória coletiva e liderança feminina na preservação do patrimônio cultural.	Manifestação Cultural; Elemento Simbólico; Gênero Musical; Celebração Religiosa; Lugar de Referência.
Texto 3	Patrimônio natural amazônico, biodiversidade, rios, ecossistemas aquáticos e importância da preservação ambiental para a manutenção da diversidade biológica.	Patrimônio Natural; Encontro das Águas dos Rios Negro e Solimões; Complexo de Conservação da Amazônia Central; Cachoeiras e Grutas do Município de Presidente Figueiredo; Serra de Curicuriari.

Fonte: elaboração própria (2026)

Os resultados indicaram diferenças na representação temática associadas à cobertura terminológica do vocabulário. Quando os conceitos centrais dos textos encontravam correspondência explícita no tesauro, foram selecionados descritores diretamente relacionados ao conteúdo, como ocorreu no texto sobre diversidade linguística indígena, representado por *Língua Indígena* e *Registro de Bens Culturais de Natureza Imaterial*. Em situações de cobertura parcial, o modelo recorreu a aproximações semânticas compatíveis com a estrutura disponível, como ocorreu no texto sobre o Maracatu. A ausência desse conceito no vocabulário levou à seleção de descritores mais amplos, como *Manifestação Cultural*, *Gênero Musical*, *Celebração*

Religiosa e *Elemento Simbólico*. Embora semanticamente coerentes com o conteúdo, esses descritores representam categorias mais gerais do que o conceito central do texto, evidenciando os limites impostos pela cobertura e pela modelagem conceitual do tesauro. Entretanto, a conformidade dos descritores com o vocabulário não implicou necessariamente pertinência temática em todas as escolhas. No Texto 1, a seleção de *Yanomami* e *Nheengatu* introduziu níveis de especificidade não explicitados no conteúdo apresentado; de modo semelhante, no Texto 3, foram selecionados descritores referentes a bens naturais específicos, embora o texto tratasse do patrimônio natural amazônico em termos gerais. Esses casos mostram que a estrutura do tesauro orientou a seleção terminológica, mas não impediu inferências expansivas do modelo, evidenciando a necessidade de distinguir conformidade terminológica com o vocabulário e pertinência temática em relação ao conteúdo do documento.

Na **terceira etapa**, a análise concentrou-se na construção iterativa das categorias utilizadas para interpretar os padrões observados nas respostas. Orientado pelo referencial da OC, o processo envolveu leitura comparativa das respostas e identificação de recorrências, deslocamentos semânticos, aproximações conceituais e formas de contextualização nas diferentes condições experimentais. Em um primeiro momento, foram identificadas cinco categorias: preservação conceitual, generalização semântica, distorção interpretativa, apagamento de contexto sociocultural e uso das relações semânticas do vocabulário. O refinamento analítico evidenciou sobreposições entre algumas dessas categorias: generalização e distorção constituíam manifestações relacionadas à preservação conceitual, enquanto o apagamento de contexto correspondia a uma manifestação da preservação ou perda do contexto sociocultural. A reorganização resultou, assim, em três dimensões analíticas complementares: preservação conceitual, preservação do contexto sociocultural e mobilização das relações semânticas do vocabulário, apresentadas no Quadro 5.

Quadro 5 – Dimensões analíticas utilizadas na interpretação dos resultados

Dimensão analítica	Descrição
Preservação conceitual	Refere-se à manutenção dos significados e características conceituais associadas aos termos analisados, permitindo observar fenômenos como preservação semântica, generalização conceitual e possíveis distorções interpretativas.
Preservação do contexto sociocultural	Refere-se à manutenção ou perda de elementos territoriais, históricos, identitários e culturais associados aos conceitos

	representados, possibilitando identificar processos de contextualização ou apagamento sociocultural.
Mobilização das relações semânticas do vocabulário	Refere-se ao uso, preservação ou reorganização das relações semânticas presentes no tesauro, incluindo relações hierárquicas, partitivas e associativas, bem como informações contextuais provenientes das notas de escopo.

Fonte: elaboração própria (2026)

As três dimensões apresentadas no quadro 5 orientaram a análise interpretativa desenvolvida na quarta etapa.

Na **quarta etapa**, a análise concentrou-se na síntese interpretativa dos resultados obtidos nas três condições experimentais. A partir das dimensões identificadas na etapa anterior: preservação conceitual, preservação do contexto sociocultural e mobilização das relações semânticas do vocabulário, foram examinadas as diferenças observadas nas respostas quando a estrutura semântica do Tesauro Yãkoana estava ou não disponível ao modelo.

1) Preservação conceitual: A análise das respostas produzidas nas três condições experimentais indica preservação dos significados centrais dos conceitos analisados, acompanhada de diferenças em sua delimitação e especificidade. Nas condições A e B, o modelo produziu interpretações compatíveis com os conceitos representados. Entretanto, enquanto na condição A as interpretações se apoiaram em associações semânticas mais amplas e em conhecimentos gerais sobre patrimônio e diversidade cultural, na condição B os conceitos foram delimitados a partir das relações hierárquicas, associativas e das notas de escopo disponibilizadas pelo Tesauro Yãkoana. Esse contraste tornou-se particularmente evidente em *Relógio Municipal*, interpretado na condição A como patrimônio urbano de forma genérica e, na condição B, como bem patrimonial específico do Centro Histórico de Manaus. Na condição C, a preservação conceitual mostrou-se mais variável. Embora tenham sido selecionados descritores correspondentes aos temas centrais dos textos, também ocorreram escolhas com níveis de generalidade ou especificidade distintos daqueles expressos no conteúdo. No texto sobre o Maracatu, a ausência do conceito no tesauro resultou no uso de categorias mais gerais; nos textos 1 e 3, por outro lado, foram selecionados descritores mais específicos do que o conteúdo apresentado. Esses resultados indicam que a estrutura do vocabulário contribuiu para delimitar o espaço terminológico da representação, mas não assegurou, por si só, correspondência entre o nível de especificidade dos descritores e o conteúdo dos documentos.

2) Preservação do contexto sociocultural: As diferenças mais expressivas entre as condições concentraram-se na contextualização sociocultural das respostas. Na condição A, as interpretações incorporaram referências à Amazônia, porém apoiadas predominantemente em conhecimentos gerais mobilizados pelo modelo. Na condição B, as respostas passaram a apresentar maior ancoragem em elementos territoriais, históricos e culturais explicitados pela estrutura semântica do tesauro. Em *Conjunto de Saberes*, por exemplo, foram incorporadas referências a sistemas agrícolas, cultivos tradicionais e práticas culturalmente situadas. De forma semelhante, conceitos como *Manifestação Cultural*, *Encontro das Águas dos Rios Negro e Solimões* e *Relógio Municipal* passaram a ser associados a elementos específicos de reconhecimento, proteção e patrimonialização presentes na estrutura disponibilizada. Na condição C, a utilização do vocabulário também introduziu elementos específicos do domínio amazônico na representação temática; contudo, essa contextualização nem sempre correspondeu ao conteúdo efetivamente expresso nos textos. A seleção de *Yanomami* e *Nheengatu* no Texto 1 e de bens naturais específicos no Texto 3 mostra que a ancoragem cultural e territorial proporcionada pelo vocabulário pode coexistir com inferências que ampliam o contexto para além daquele explicitado no documento.

3) Mobilização das relações semânticas do vocabulário: A terceira dimensão indicou que o modelo não apenas utilizou os termos disponibilizados pelo tesauro, mas incorporou à interpretação relações e informações contextuais presentes em sua estrutura. Relações hierárquicas (TG), associativas (TR) e partitivas (TGP), bem como informações contidas nas notas de escopo, foram mobilizadas para delimitar conceitos, estabelecer vínculos e contextualizar os bens representados. Em *Relógio Municipal*, por exemplo, a relação partitiva com o *Centro Histórico de Manaus* permitiu situar o bem em um conjunto patrimonial mais amplo, enquanto, em *Encontro das Águas dos Rios Negro e Solimões*, as relações e informações contextuais disponibilizadas pelo tesauro orientaram a interpretação para aspectos relacionados ao reconhecimento e à proteção patrimonial. Na condição C, a estrutura do vocabulário também serviu de referência para a seleção e a justificativa dos descritores, indicando que o tesauro foi utilizado não apenas como fonte terminológica, mas como estrutura de orientação semântica para a representação temática. Os casos de seleção de descritores mais específicos do que o conteúdo dos textos, contudo, mostram que a mobilização dessas relações não assegurou, por si só, a pertinência temática das escolhas realizadas pelo modelo. Esses resultados convergem com a perspectiva de Dahlberg (1993) sobre a centralidade das relações conceituais na OC e

dialogam com as discussões de Beghtol (1995; 2002) e Barité (2009; 2011; 2022) acerca da importância do contexto cultural na construção e utilização de sistemas de representação.

Consideradas em conjunto, as três dimensões mostram que a disponibilização da estrutura semântica do Tesauro Yãkoana esteve associada a interpretações mais delimitadas e contextualizadas e forneceu referências terminológicas e relacionais para a representação temática. Os resultados também evidenciaram, contudo, que essa ancoragem não elimina inferências produzidas pelo modelo nem assegura, por si só, a pertinência temática dos descritores selecionados. O papel do vocabulário controlado apresenta-se, assim, menos como mecanismo de determinação das respostas do que como estrutura de mediação semântica que orienta e delimita o espaço de interpretação e representação.

5 CONSIDERAÇÕES FINAIS

Os resultados indicam diferenças nos processos de interpretação conceitual e representação temática quando a estrutura semântica do Tesauro Yãkoana estava disponível ao modelo. Nas tarefas de interpretação, as respostas produzidas com acesso ao vocabulário apresentaram maior delimitação conceitual e ancoragem em elementos territoriais, históricos e culturais específicos do domínio amazônico. Nas tarefas de representação temática, a estrutura do tesauro orientou a seleção dos descritores e forneceu relações semânticas e informações contextuais para a justificativa das escolhas.

A análise também evidenciou, contudo, limites dessa mediação. A correspondência dos descritores com o vocabulário não assegurou, por si só, sua pertinência temática em relação ao conteúdo dos documentos. Quando o conceito central não estava explicitamente representado no tesauro, como no caso do Maracatu, o modelo recorreu a categorias mais gerais disponíveis no vocabulário. Em outros casos, selecionou descritores mais específicos do que o conteúdo apresentado. Esses resultados mostram que a cobertura e a modelagem conceitual do SOC delimitam o espaço terminológico disponível para a representação, mas não eliminam inferências expansivas produzidas pelo modelo.

Do ponto de vista da OC, os resultados sugerem que tesauros e outros SOC podem atuar como estruturas de mediação semântica em processos automatizados de interpretação e representação em ambientes de IA, fornecendo ancoragem conceitual, relacional e contextual sem determinar integralmente as respostas produzidas. Essa constatação reforça a importância

de distinguir conformidade terminológica com o vocabulário e pertinência temática em relação ao documento, especialmente em domínios culturalmente situados.

Entre as limitações do estudo, destacam-se o número reduzido de termos e textos analisados, a utilização de um único modelo de linguagem, a realização de uma única execução para cada interação e a investigação de um único domínio cultural. Além disso, as condições A e B utilizaram instruções distintas, adequadas às informações disponibilizadas em cada situação, de modo que as diferenças observadas devem ser interpretadas qualitativamente, e não como efeito causal isolado da disponibilização do tesauro. Essas limitações delimitam o alcance dos achados e reforçam o caráter exploratório do estudo.

Como perspectivas futuras, recomenda-se ampliar o experimento para outros SOC, domínios culturais e modelos de linguagem, incorporar múltiplas execuções e procedimentos de validação humana da pertinência dos descritores e investigar o potencial dos vocabulários controlados como mecanismos de orientação e governança semântica em contextos de IA generativa. Tais investigações poderão contribuir para compreender em que condições estruturas semânticas especializadas favorecem interpretações e representações simultaneamente coerentes com o domínio, culturalmente contextualizadas e pertinentes ao conteúdo documental.

Em conjunto, os resultados evidenciam a contribuição potencial da OC para o desenvolvimento e uso de estruturas semânticas capazes de orientar processos de interpretação e representação em ambientes de IA, ao mesmo tempo que ressaltam a permanência da necessidade de validação das inferências produzidas pelos modelos.

REFERÊNCIAS

ATELIÊ AMAZÔNICO. **Ateliê Amazônico**: tradição e contemporaneidade. Programa de Educação Tutorial de Comunicação da Universidade Federal do Amazonas (PETCOM/UFAM), [2026?]. Disponível em: https://atelieamazonico.weebly.com/. Acesso em: 12 jul. 2026.

BARITÉ, Mario. Cultural warrant as a justification in knowledge organization systems: a critical approach. **Palabra Clave**, Buenos Aires, v. 1, n. 1, p. 2-11, 2011. Disponível em: https://www.palabraclave.fahce.unlp.edu.ar/article/view/PCv1n1a02/0. Acesso em: 12 jul. 2026.

BARITÉ, Mario. Garantía literaria y normas para construcción de vocabularios controlados. **Scire**: representación y organización del conocimiento, Zaragoza, v. 15, n. 2, 2009. Disponível em: https://ibersid.eu/ojs/index.php/scire/article/view/3709. Acesso em: 12 jul. 2026.

BARITÉ, Mario. Las garantías más allá del horizonte de la organización del conocimiento: proyecciones posibles. *In*: BARROS, Thiago Henrique Bragato; LAIPELT, Rita do Carmo Ferreira (org.). **Organização e Representação do Conhecimento em Múltiplas Abordagens**. São Paulo: Pimenta Cultural, 2022. p. 124-148.

BEGHTOL, Clare. A proposed ethical warrant for global knowledge representation and organization systems. **Journal of Documentation**, [*s. l.*], v. 58, n. 5, p. 507-532, 2002. Disponível em: https://www.emerald.com/jd/article-abstract/58/5/507/203595/A-proposed-ethical-warrant-for-global-knowledge?redirectedFrom=fulltext. Acesso em: 12 jul. 2026.

BEGHTOL, Clare. Domain analysis, literary warrant, and consensus: the case of fiction studies. **Journal of the American Society for Information Science**, Hoboken, v. 46, n. 1, p. 30-44, 1995. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-4571(199501)46:1%3C30::AID-ASI4%3E3.0.CO;2-F. Acesso em: 12 jul. 2026.

BEGHTOL, Clare. Semantic validity: concepts of warrant in bibliographic classification systems. **Library Resources & Technical Services**, [*s. l.*], v. 30, n. 2, p. 109-125, 1986. Disponível em: https://alair.ala.org/server/api/core/bitstreams/6ff3860c-d197-4fbf-9cb4-2b0d9e0be142/content. Acesso em: 12 jul. 2026.

BRAUN, Virginia; CLARKE, Victoria. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. **Counselling and Psychotherapy Research**, [*s. l.*], v. 21, n. 1, p. 37-47, 2021. Disponível em: https://onlinelibrary.wiley.com/doi/10.1002/capr.12360. Acesso em: 12 jul. 2026.

BRAUN, Virginia; CLARKE, Victoria. Reflecting on reflexive thematic analysis. **Qualitative Research in Sport, Exercise and Health**, [*s. l.*], v. 11, n. 4, p. 589-597, 2019. Disponível em: https://www.tandfonline.com/doi/full/10.1080/2159676X.2019.1628806. Acesso em: 12 jul. 2026.

DAHLBERG, Ingetraut. Knowledge organization: its scope and possibilities. **Knowledge Organization**, Frankfurt, v. 20, n. 4, p. 211-222, 1993. Disponível em: https://www.imrpress.com/journal/KO/20/4/10.5771/0943-7444-1993-4-211. Acesso em: 12 jul. 2026.

DRABINSKI, Emily. Teaching about class in the library. **Radical Teacher**, [*s. l.*], n. 85, p. 15-16, Fall 2009. Disponível em: https://paperzz.com/doc/7100431/teaching-about-class-in-the-library. Acesso em: 12 jul. 2026.

FREITAG, Raquel Meister Ko. et al. Brazilian linguistic diversity platform: linguistic data for a Brazilian AI. **Cadernos de Linguística**, Campinas, v. 6, n. 4, e863, 2025. Disponível em: https://cadernos.abralin.org/index.php/cadernos/article/view/863. Acesso em: 12 jul. 2026.

GALLEGOS, Isabel O.; ROSSI, Ryan A.; BARROW, Joe; TANJIM, Md Mehrab; KIM, Sungchul, DERNONCOURT, Franck; YU, Tong; ZHANG, Ruiyi; AHMED, Nesreen K. Bias and fairness in large language models: a survey. **Computational Linguistics**, Cambridge, v. 50, n. 3, p. 1097-1179, 2024. Disponível em: https://direct.mit.edu/coli/article/50/3/1097/121961/Bias-and-Fairness-in-Large-Language-Models-A. Acesso em: 12 jul. 2026.

HJØRLAND, Birger. What is knowledge organization (KO)?. **Knowledge Organization**, Frankfurt, v. 35, n. 2-3, p. 86-101, 2008. Disponível em: https://storage.imrpress.com/imr/journal/KO/article/505437/1752847178259.pdf. Acesso em: 12 jul. 2026.

HULME, Edward Wyndham. Principles of Book Classification. **Library Association Record**, Londres, v. 13, p. 354-358, 1911.

OLSON, Hope A. **The power to name**: locating the limits of subject representation in libraries. Dordrecht: Kluwer Academic Publishers, 2002. Disponível em: https://link.springer.com/book/10.1007/978-94-017-3435-6. Acesso em: 12 jul. 2026.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial intelligence:** a modern approach. 4. ed. Harlow: Pearson Education Limited, 2022.

SPECIA, Lucia; RINO, Lucia Helena Machado. **Representação semântica:** alguns modelos ilustrativos. São Carlos: Núcleo Interinstitucional de Linguística Computacional (NILC), 2002. (Série de Relatórios do NILC).

TRINDADE, Thais Lima. **Yãkoana Tesauro dos Bens Culturais do Amazonas:** organização e recuperação da informação do patrimônio cultural amazonense. 2025. 353 f. Tese (Doutorado em Ciência da Informação) — Universidade de Brasília, Brasília, 2025. Disponível em: http://repositorio.unb.br/handle/10482/52570. Acesso em: 12 jul. 2026.

UNESCO. **Recomendação sobre a ética da inteligência artificial**. Paris: UNESCO; Representação da UNESCO no Brasil, 2022.

AGRADECIMENTOS

Agradecimento ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelas bolsas de produtividade – PQ (processos 307765/2023-7; 313645/2020-5; 306713/2025-0).