



Compreendendo thumbnails: análise discursiva e visual no youtube

Understanding thumbnails: discursive and visual analysis on youtube

Ulisses Tadeu Vaz De Oliveira ^(a); Geovana Pin Serion ^(b)

^a Universidade Federal do Mato Grosso do Sul (UFMS), Brasil – ulisvaz@gmail.com

^b Universidade Federal do Mato Grosso do Sul (UFMS), Brasil – geovana.serion@gmail.com

Resumo: O avanço tecnológico possibilitou o surgimento das redes sociais e de plataformas digitais de compartilhamento de vídeos, como o YouTube, que deram origem a novos gêneros textuais adaptados ao ambiente online. Entre eles, destaca-se a thumbnail, elemento visual estratégico cuja função é atrair a atenção do público e aumentar a visibilidade dos vídeos. Em conjunto com os títulos, as thumbnails apresentam características linguísticas e visuais próprias, potencializadas pelas tecnologias de edição e design gráfico, o que as torna centrais na dinâmica comunicativa do YouTube. Este estudo tem como objetivo analisar aspectos composicionais e discursivos do gênero thumbnail, com ênfase em seu potencial persuasivo. Para tanto, adota-se como base teórica a Linguística Sistêmico-Funcional, a Gramática do Design Visual, a Linguística da Internet, a abordagem da Multimodalidade e o modelo de Estrutura Potencial Genérica. O corpus é composto por 45 thumbnails e títulos, coletados durante o mês de maio de 2025 na seção “em alta” da plataforma. A análise investiga como esses elementos se constituem como gêneros multimodais, os padrões linguísticos e visuais predominantes e a função comunicativa do gênero. Os resultados indicam a predominância de títulos com construções linguísticas específicas para diferentes categorias de vídeos, aliados a thumbnails com imagens centralizadas e forte contraste cromático, maximizando a visibilidade e o apelo ao clique. Também foi observada a presença recorrente de apelos emocionais e de recursos multimodais que reforçam estratégias de bait, como a promessa de novidade, a exclusividade e a criação de expectativa.

Palavras-chave: thumbnails; Linguística Sistêmico-Funcional; Gramática do Design Visual; Multimodalidade

Abstract: Technological advances have led to the emergence of social media and digital video-sharing platforms such as YouTube, giving rise to new textual genres adapted to the online environment. Among them, thumbnails stand out as strategic visual elements designed to attract viewers' attention and increase video visibility. Together with video titles, thumbnails display distinctive linguistic and visual features, enhanced by editing tools and graphic design technologies, making them central to YouTube's communicative dynamics. This study aims to analyze the compositional and discursive aspects of the thumbnail genre, with emphasis on its

persuasive potential. The research is theoretically grounded in Systemic Functional Linguistics, Visual Design Grammar, Internet Linguistics, Multimodality, and Generic Structure Potential. The corpus consists of 45 thumbnails and titles collected in May 2025 from YouTube's "trending" section. The analysis investigates how these elements function as multimodal genres, identifies predominant linguistic and visual patterns, and explores their communicative function. Findings indicate a prevalence of titles with specific linguistic constructions across different video categories, combined with thumbnails featuring centralized images and strong color contrast, maximizing visibility and click appeal. Emotional appeals and multimodal resources reinforcing bait strategies—such as promises of novelty, exclusivity, and curiosity – were also frequently observed.

Keywords: thumbnails; Systemic Functional Linguistics; Visual Design Grammar; Multimodality

Introdução

A linguagem permeia diversos aspectos da vida humana e, desde o surgimento da internet, a multiculturalidade e a globalização têm afetado decisivamente os fenômenos linguísticos. As transformações para a comunicação humana têm sido particularmente impulsionadas pelas redes sociais e plataformas de compartilhamento de vídeos, como o *YouTube*. Nesse ambiente digital, surgem novos gêneros textuais que, adaptados ao contexto online, refletem e atendem às demandas dos consumidores desse tipo de conteúdo.

Ciências como a Linguística da Internet (CRYSTAL, 2011) têm chamado a atenção para fenômenos linguísticos na internet e os gêneros digitais. Um exemplo desses gêneros, presente em nosso dia a dia, mas quase imperceptível, é a *thumbnail*.

Nas novas dinâmicas sociais digitais, a *thumbnail* vem a equivaler ao papel da “capa” ou “primeira página” de veículos jornalísticos de tradição impressa (e.g. revista e jornal). Nas páginas online, este gênero se apresenta associado aos títulos dos vídeos, desempenhando um importante papel comunicativo na captura da atenção do público e na amplificação da visibilidade dos vídeos. Esses elementos são feitos para

atrair cliques e engajamento dos usuários e apresentam características híbridas, incorporando tanto traços de gêneros tradicionais, como anúncios publicitários e manchetes de notícias, quanto inovações permitidas pelas tecnologias contemporâneas de design gráfico e edição. Diante desse cenário, este artigo tem como objetivo analisar aspectos composicionais e discursivos do gênero *thumbnail*, com ênfase na persuasão, em 45 *thumbnails* e títulos (capturas), selecionados durante o mês de maio de 2025, conforme surgiam na parte “em alta” do *YouTube*, exceto a categoria “podcasts” que foram selecionados os últimos vídeos postados no dia 01/06/2025 dos canais com mais inscritos. Desse modo, pretende-se identificar padrões linguísticos e visuais predominantes, além de compreender a função comunicativa desses elementos nos segmentos analisados – i.e. em particular, em vídeos de canais brasileiros que atuam nos nichos/categorias: *entretenimento*, *games*, *reação/comentário*, *música*, *vlog*, *jornalismo* (*notícia*, *entrevista*), *esporte*, *podcast* e *educação/curiosidades*. Tal investigação visa a compreender como essas ferramentas textuais e visuais influenciam a experiência do usuário e afetam as práticas de consumo de conteúdo na sociedade digital.

A relevância desta pesquisa está primariamente baseada na sua contribuição para o estudo dos gêneros multimodais no ambiente digital, além de ampliar as discussões sobre práticas discursivas em plataformas digitais. Ao atender tal orientação, o estudo traz à luz um gênero que, apesar de desprestigiado ou ignorado, faz parte da nossa rotina de forma decisiva. Adicionalmente, as reflexões críticas sobre a função persuasiva e os impactos sociais dos *thumbnails* na comunicação online destacam o seu papel e potencial de engajamento e manipulação junto aos usuários de plataformas digitais.

Para a realização deste estudo, adota-se um arcabouço teórico que integra diferentes abordagens, permitindo uma análise abrangente das *thumbnails* e títulos de vídeos como gêneros multimodais no *YouTube*.

Primeiramente, a Linguística Sistêmico-Funcional – LSF (HALLIDAY, 1994) é utilizada para explorar os aspectos linguísticos desses elementos, considerando suas metafunções ideacional, interpessoal e textual para compreender como representam o conteúdo, interagem com o público e estruturam as mensagens das *thumbnails*. Associada à LSF, a Gramática do Design Visual – GDV, de Kress e van Leeuwen (2006) permite analisar os recursos visuais, como a composição, cores, tipografia e outros elementos gráficos, desses textos.

Nesta pesquisa, mobilizamos também os fundamentos teóricos da persuasão em articulação com a Análise do Discurso Crítica (ADC) e com a Teoria da Avaliatividade (MARTIN; WHITE, 2005), no âmbito da Linguística Sistêmico-Funcional. Conceitos como sedução e convicção permitem compreender as estratégias discursivas acionadas nas *thumbnails* e títulos, sobretudo por meio dos recursos de atitude, comprometimento e gradação. Essa abordagem evidencia como enunciados verbais e visuais atraem a atenção do público enquanto configuram relações de poder e posicionamento ideológico no ambiente digital.

De forma integrada, o estudo também recorre à Linguística da Internet (CRYSTAL, 2011), à Multimodalidade e à Estrutura Potencial Genérica (HASAN, 1984), possibilitando articular os modos verbal e visual de modo a identificar estruturas recorrentes e funções comunicativas próprias do gênero. Assim, todas essas perspectivas se complementam na investigação da composição formal das *thumbnails* e títulos e de suas implicações sociais, discursivas e práticas no ecossistema digital. O corpus

é formado por cinco capturas de vídeos/títulos de canais brasileiros do *YouTube*, representativos de diferentes nichos de grande audiência.

Os resultados do estudo demonstraram a prevalência de padrões linguísticos específicos nos títulos para diferentes categorias. Em paralelo, as *thumbnails*, como demonstraremos, apresentaram aspectos composicionais regulares (e.g., imagens centralizadas e forte contraste cromático) para maximizar visibilidade e persuasão. Observou-se também a presença recorrente de apelos emocionais e de estruturas multimodais que reforçam diferentes tipos de *bait*, os quais, tal qual o bojo teórico do estudo, são discriminados na seção teórica que se segue.

2. O que são *thumbnails*?

A rápida evolução das tecnologias digitais tem promovido transformações indelévels nas interações sociais e nas estruturas comerciais e culturais contemporâneas. Esse fenômeno modifica a forma como empresas e criadores de conteúdo se comunicam com os consumidores e também redefine o papel dos indivíduos como participantes ativos em uma sociedade digital interativa.

Segundo Ceci (2025), é nesse contexto de transformação que surge o *YouTube*, em 2005, sendo adquirido pela empresa Google em 2006. Hoje, a plataforma possui um valor de marca estimado em 31,7 bilhões de dólares americanos e, em 2024, contabilizou mais de 2,5 bilhões de espectadores globais – sendo 144 milhões apenas no Brasil. O *YouTube*, além de representar uma das principais plataformas digitais contemporâneas, se consolidou como um espaço fértil para a expressão criativa (música, entretenimento, arte) e para o uso de estratégias de marketing digital.

Para Godin (2005), o marketing não é mais sobre as coisas que você faz, mas sobre as histórias que você conta. Nesse cenário, o fenômeno dos vídeos online transformou-se num mercado multimilionário, o qual vem redefinindo as dinâmicas de engajamento entre produtores de conteúdo e audiência. Em meio à intensa competição por atenção, os criadores precisam desenvolver recursos visuais que não apenas informem, mas que atraiam o olhar e despertem interesse imediato.

Com base nessa finalidade, surgem as *thumbnails*, ou miniaturas, que funcionam como a primeira interface visual entre o conteúdo e o usuário. De acordo com Jang et al. (2023, p. 8171), trata-se de “[...] uma imagem criada resumindo o conteúdo do vídeo e adicionando efeitos visuais [...] É uma imagem que permite aos usuários obter informações básicas sobre o vídeo antes de selecionar o conteúdo” (tradução nossa). Em outras palavras, a thumbnail atua como um convite visual: sua atratividade pode determinar se um vídeo será assistido ou ignorado. Portanto, além de terem uma função informacional – antevêem o conteúdo do vídeo, têm também função persuasiva – atrair cliques, transmitir emoção instantaneamente, gerar engajamento.

A importância da *thumbnail* não deve ser menosprezada, inclusive sob a perspectiva de sua construção estética. Embora inseridas em contextos histórico-discursivos profundamente distintos, as miniaturas medievais (também conhecidas como iluminuras) e as thumbnails digitais podem ser comparadas quanto à sua função visual condensada e estratégica. As miniaturas, aplicadas à margem dos manuscritos medievais, tinham como objetivo ampliar a inteligibilidade do texto sagrado ou narrativo, ilustrando e destacando trechos-chave por meio de composições visualmente marcadas (HOEFNAGEL, 1570; TURNER, 2001).

De maneira análoga às iluminuras, as thumbnails, imagens reduzidas associadas a vídeos em plataformas como o *YouTube*, desempenham um importante papel na comunicação digital contemporânea. Enquanto as miniaturas operavam predominantemente em uma lógica de mediação religiosa e estética, as thumbnails atuam sob um regime persuasivo e performativo ao articular elementos do design gráfico e da retórica digital para competir por visualizações em ambientes altamente saturados.



Fig. 1 – Exemplo de iluminura medieval

Fonte: Iluminura da letra “B”. Fólio 7r, Saltério “St Omer” ou “Rede”, MS 98, Christ Church, Universidade de Oxford, 1201–1300. Disponível em:

<https://digital.bodleian.ox.ac.uk/objects/1159e59e-b06e-44d9-bb0a-b5127a5b36fe/surfaces/9da1b4c2-2c04-40e9-a171-dc4b656af141>

Miniaturas e *thumbnails* partilham uma função multimodal de condensação significativa, ainda que inseridas em regimes semióticos e pragmáticos distintos. O estudo de Jang et al. (2023) identificou que atributos das *thumbnails*, como o brilho, a cor e a simplicidade visual, são fatores decisivos na sua eficácia. Imagens mais claras e coloridas tendem a capturar a atenção com maior facilidade do que aquelas com paletas apagadas, e a clareza composicional contribui para a compreensão imediata da mensagem visual. Jang et al. (2023) também enfatizam que o

uso de design harmônico e a presença de expressões faciais humanizadas podem estabelecer conexões emocionais com os usuários, incentivando o clique.

Apesar de ser um gênero emergente, a *thumbnail* compartilha diversas características com gêneros consolidados associados ao hipergênero publicitário. Para Lara e Souza (2019), as *thumbnails* “tem por finalidade a divulgação de um produto, um serviço, uma marca, com fins comerciais. Assim, sua função primeira é fazer escoar as mercadorias, convencer as pessoas a adquirirem o que foi produzido”. De modo semelhante, ao publicar um vídeo, criadores de conteúdo devem produzir uma *thumbnail* chamativa, capaz de antecipar e representar o conteúdo do vídeo numa imagem, atraindo cliques e visualizações – que são, por sua vez, convertidos em visibilidade, engajamento e oportunidades de monetização.

Com uma finalidade discursiva bem definida, o presente estudo interessa-se na descrição dos aspectos que, combinados, estabelecem este gênero emergente e tão presente em nossas vidas. Para tal, nas seguintes seções, discutimos o panorama teórico que sustenta nossa análise.

3. Gênero e a multimodalidade dos gêneros digitais

As *thumbnails*, amplamente utilizadas em plataformas digitais, como explicamos, configuram-se como uma manifestação contemporânea de gêneros textuais visuais. Essas pequenas imagens cumprem funções estéticas e também possuem características funcionais e comunicativas claras: sintetizam e antecipam o conteúdo que representam e incentivam a interação do usuário. Dessa forma, é possível analisá-las sob a ótica de gêneros textuais digitais por sua estrutura recorrente e adaptável aos contextos digitais. Para Marcuschi (2008, p. 20),

Toda linguagem é dialógica, ou seja, todo enunciado é sempre um enunciado de alguém para alguém. [...] Daí a noção de gênero como enunciado responsivo "relativamente estável", o que está de acordo com a ideia de linguagem como atividade interativa e não como forma ou sistema.

A linguagem, enquanto produto da comunicação, manifesta-se em diferentes semioses, como imagens, gestos, sons e textos escritos. Entre essas formas, a língua é um dos principais meios de significação e se realiza socialmente por meio dos gêneros textuais, os quais atendem a demandas comunicativas de diversos contextos sociais. De acordo com Martin, gênero é "uma atividade estruturada, orientada por objetivos e com propósito, na qual os falantes se envolvem como membros de nossa cultura" (1997, p.13 – tradução livre)¹. Essa perspectiva destaca a natureza funcional dos gêneros na organização e comunicação dentro de contextos sociais específicos,

De acordo com Karwoski e Gaydeczka:

[...] o processo de produção textual não é mais exclusivamente linguístico, integra imagem, som, movimento; além disso, não se vivencia mais uma produção estritamente individual ou de mão única (aluno-professor), mas colaborativa – mais de um sujeito contribui para a produção e retextualização" (KARWOSKI; GAYDECZKA, 2013, p. 5).

Ou seja, como os gêneros escritos e orais, as *thumbnails* também são organizadas em estágios, cada um desempenhando uma função específica para atingir finalidades discursivas determinadas. Contudo, sua estrutura distingue-se por estar enraizada na verbo-visualidade, integrando elementos textuais e visuais que colaboram para a construção de sentidos e a interação com o público. Essa característica multimodal alinha-se às

¹ No original: "a staged, goal-oriented, purposeful activity in which speakers engage as members of our culture."

demandas comunicativas do meio digital e dos gêneros textuais emergentes.

As tecnologias digitais impulsionaram a multimodalidade por proporcionar uma miríade de aplicações e funcionalidades associadas às diversas semioses que canalizam tais gêneros. Como observado por Halliday e Hasan (1989) e recuperado por Isola (2012), a construção de gêneros textuais envolve inúmeros recursos semióticos verbais e não verbais. Essa combinação é destaque nas plataformas digitais, devido à ausência de barreiras físicas e ao acesso quase ilimitado às ferramentas de criação.

A evolução das tecnologias e implementação de funcionalidades como, no contexto atual, a inteligência artificial, afetam diretamente a constituição dos gêneros digitais. Este fenômeno é exemplificado por Marcuschi (2002), destacando a comunicação social e em determinados momentos históricos. Exemplos desse fenômeno incluem a transição do telegrama para mensagens instantâneas, ou a evolução de entrevistas radiofônicas para podcasts digitais, que integram características de gêneros anteriores a novos formatos interativos, como transmissões ao vivo e chats.

4. Estrutura Potencial Genérica

A crescente utilização de *thumbnails* nas plataformas digitais, como elemento da comunicação visual e textual, levanta questões sobre sua caracterização enquanto gênero textual e sua organização discursiva. Enquanto gêneros tradicionais, como artigos, cartas ou entrevistas, apresentam estruturas mais explicitamente reconhecíveis, as *thumbnails* desafiam essas convenções ao unirem texto e imagem em um espaço reduzido e de alto impacto visual.

Neste artigo, mobiliza-se o arcabouço teórico da Estrutura Potencial Genérica (EPG), definida por Hasan (1989) como a organização funcional dos elementos que configuram um gênero. A EPG vai além da estrutura verbal, abrangendo toda a linguagem significativa do texto, inclusive sua dimensão multimodal. Essa configuração considera o ambiente de produção e interpretação, incluindo autor, leitor e contexto sociocultural (SIMÕES, 2021). O Quadro 1 apresenta seus conceitos básicos.

Composição da EPG	Descrição	Perguntas para a investigação
Estágios OBRIGATÓRIOS	Estes são componentes essenciais que devem ser incluídos em qualquer ocorrência do gênero.	O que sempre se repete no corpus?
Estágios OPCIONAIS	Estes componentes podem aparecer ou não, dependendo da intenção do autor e das características particulares da obra.	O que surge somente algumas vezes?
Estágios SEQUENCIAIS	Relacionam-se à ordem em que os estágios aparecem; indicam que há uma sequência típica esperada entre os componentes do gênero.	Em que ordem os estágios costumam aparecer?
Estágios ITERATIVOS (ou recursivos)	Relacionam-se a componentes que podem ser repetidos ou que aparecem em múltiplas instâncias dentro do gênero.	O que se repete mais de uma vez em diversas posições no mesmo exemplo do corpus?

Quadro 1 – Estágios na Estrutura Potencial Genérica
Fonte: Autores, com base em Hasan (1989)

Segundo Hasan (1989), o potencial genérico está relacionado às diversas possibilidades de organização de um gênero específico. Os gêneros, nesse caso, possuem uma estabilidade relativa, garantida pelos elementos obrigatórios que estão sempre presentes em qualquer ocorrência do gênero e por elementos opcionais e recursivos, que ficam à critério do criador do texto colocá-los ou não, uma vez que a seleção desses elementos para formar a estrutura do texto é semelhante às decisões tomadas no sistema linguístico, pois ambas são influenciadas pelo

contexto (AUTOR 1, p.72). Ao entender *thumbnails* como um gênero que surgiu e se desenvolveu no contexto da internet e da multimodalidade, a EPG vai permitir que se analise a organização interna desse gênero.

5. A Linguística Sistêmico-Funcional e a Gramática do Design Visual

A Linguística Sistêmico-Funcional (LSF), desenvolvida por Michael Alexander Kirkwood Halliday, tem como base a concepção de que a linguagem é, antes de tudo, um fenômeno social voltado à comunicação. Consolidada especialmente a partir da obra *An Introduction to Functional Grammar* (1994), a teoria parte do princípio de que o significado linguístico decorre das escolhas feitas pelos usuários em contextos específicos. Ao adotar uma perspectiva paradigmática, Halliday argumenta que as estruturas da língua não são fixas, mas moldadas pelas opções significativas que os falantes realizam, influenciando diretamente a forma e a função das expressões linguísticas.

A linguagem, sob a perspectiva da Linguística Sistêmico-Funcional, é concebida como um sistema de significados que evoluiu primariamente para cumprir funções sociais. A funcionalidade é, portanto, o princípio organizador da linguagem, o que implica dizer que sua forma está diretamente relacionada às finalidades que ela busca atender. Halliday e Matthiessen (2014) identificam três metafunções principais que coexistem em qualquer uso da língua:

- a) A metafunção ideacional, responsável por representar a experiência humana e os processos do mundo, associada ao componente campo do contexto;
- b) A metafunção interpessoal, que regula as relações sociais entre os interlocutores, vinculada à variável *Tenor* (Relações);

c) A metafunção textual, que organiza o discurso de forma coesa e coerente, conectando-o à situação de uso, relacionada ao parâmetro modo.

Dessa forma, compreende-se que o uso da linguagem está indissociavelmente ligado a fatores contextuais, sociais e interacionais, e que os significados produzidos na língua só podem ser plenamente compreendidos à luz de seu contexto de uso.

A partir dessa base teórica, a LSF oferece instrumentos bem consolidados para a análise de textos verbais em sua complexidade funcional e contextual. No entanto, considerando os gêneros digitais contemporâneos, como é o caso das *thumbnails*, é necessário expandir o olhar para além da linguagem verbal, incorporando outras formas semióticas de significação. Nesse sentido, a Gramática do Design Visual (GDV), proposta por Kress e van Leeuwen (2006), surge como uma extensão da LSF no domínio do visual e fornece ferramentas para que se compreenda como imagens também realizam funções comunicativas de maneira sistemática e organizada.

6. A Gramática do Design Visual

Assim como a LSF, a GDV parte do princípio de que os elementos visuais não são aleatórios, mas estruturados segundo princípios funcionais. A GDV retoma as três metafunções da LSF – ideacional, interpessoal e textual – e as aplica à análise visual: como os textos visuais representam o mundo, constroem relações e organizam-se de modo coerente. No caso das *thumbnails*, essas três funções podem ser observadas, por exemplo, na escolha dos elementos imagéticos (metafunção ideacional), na construção de apelos ao público e expressões faciais dos personagens (metafunção

interpessoal), e na composição gráfica que direciona o olhar do espectador (metafunção textual).

Na presente pesquisa, a integração entre LSF e GDV para o estudo de gêneros multimodais, como as *thumbnails*, articula recursos verbais e visuais em um mesmo espaço comunicativo. Ao combinar essas abordagens, o impacto e a eficácia comunicativa do gênero podem ser analisados e descritos em termos de seus aspectos centrais no ecossistema digital atual, sobretudo em plataformas como o *YouTube*, onde a concorrência pela atenção do usuário é intensa e a comunicação se dá em milésimos de segundo.

Neste artigo, a GDV será utilizada como eixo principal da análise, uma vez que as *thumbnails* são essencialmente visuais e gráficas, ainda que frequentemente associadas a elementos verbais como títulos e textos. A partir da articulação entre as metafunções e os princípios composicionais definidos por Kress e van Leeuwen (2006), busca-se identificar padrões de construção visual que caracterizam esse gênero, bem como compreender suas estratégias de persuasão, visibilidade e interação com o público.

Kress e van Leeuwen (2006) argumentam que os significados visuais são realizados de maneira distinta dos verbais, mas igualmente regidos por normas socioculturais. Cores, composição, ângulos de visão, enquadramentos e distâncias sociais são exemplos de recursos que, longe de serem neutros, refletem escolhas e convenções que circulam nas práticas sociais. Dessa forma, a GDV propõe a análise dos textos imagéticos a partir de três funções principais: representacional, interativa e composicional, as quais correspondem, respectivamente, às metafunções ideacional, interpessoal e textual da LSF.

6.1 Função Representacional: Como as imagens representam o mundo

A função representacional diz respeito à maneira como os elementos visuais representam eventos, ações, entidades e relações no mundo. Na GDV, essa função é dividida entre representações narrativas e conceituais. As narrativas envolvem ação, movimento ou transformação, enquanto as conceituais referem-se a classificações, identificações e atributos.

No contexto das *thumbnails*, a função representacional pode ser observada nos processos visuais que indicam ações (como expressões faciais exageradas, gestos, vetores direcionais), ou que posicionam os participantes em situações específicas. As imagens podem representar, por exemplo, um influenciador em ato de surpresa (vetor gestual), ou a associação de elementos icônicos (como emojis ou setas) para sinalizar movimento, choque ou mistério.

Na representação narrativa, quando participantes estão conectados por um vetor, eles são representados como fazendo algo para ou entre um ao outro. Esses padrões vetoriais são a narrativa. Nas figuras, a seguir, são apresentados vetores (setas em vermelho), indicando processos, nas principais modalidades em que ocorrem:

- Transacional: representa ações ou interações entre participantes dentro de uma imagem (Figuras 2 e 3).

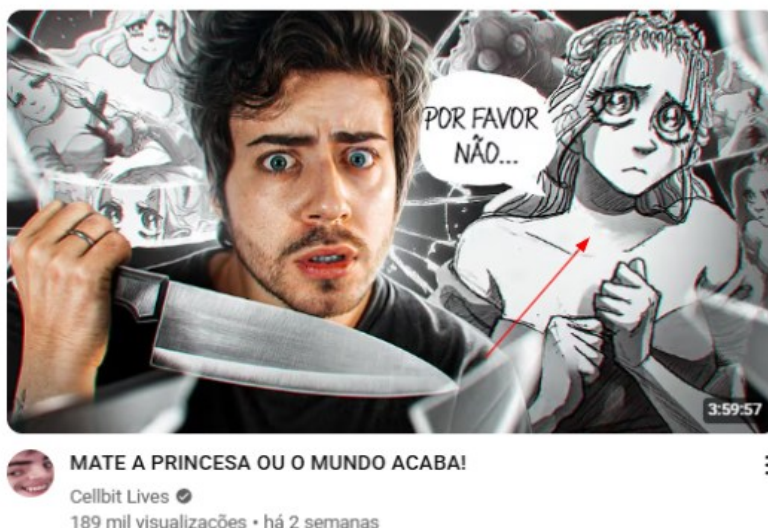


Fig. 2 – Processo Transacional (exemplo 1)

Fonte: Cellbit Lives, YouTube. Disponível em

<https://www.youtube.com/watch?v=4pmRioSAV2s>. Acesso em ago/2025.



Fig. 3 – Processo Transacional (exemplo 2)

Fonte: orochidois, YouTube. Disponível em

<https://www.youtube.com/watch?v=arBqN0n3Qo8>. Acesso em ago/2025.

- Reacional: indica uma reação ou olhar entre participantes, geralmente envolvendo uma relação de olhar. Podem ser de dois tipos: (I) reação transacional – realizada por uma linha ocular, por

um olhar direcional do(s) Participante(s), que são, neste caso, chamados de *reator(es)*. Os participantes que são observados são *fenômenos*; e (II) reação não-transacional, que envolve apenas um participante sem uma interação direta com outro participante (Figuras 4 e 5).



Fig. 4 – Processos Reacionais – Reação Transacional

Fonte: Felipe Neto, *YouTube*. Disponível em <https://www.youtube.com/watch?v=pqHgRtlaop4>. Acesso em ago/2025.



Fig. 5 – Processos Reacionais – Reação Não-Transacional

Fonte: República Coisa de Nerd, *YouTube*. Disponível em <https://www.youtube.com/watch?v=7uqkiUVE8Pl>. Acesso em ago/2025.

- Mental: indica processos mentais ou emocionais entre participantes, muitas vezes envolvendo olhares ou gestos que sugerem pensamento ou sentimento.
- Circunstancial: relaciona-se às circunstâncias ou contexto em que a ação ocorre, sem foco em uma interação direta entre participantes.

Essas escolhas visuais constroem representações que interagem diretamente com a experiência do público-alvo, especialmente em gêneros digitais como os vídeos no *YouTube*.

6.2. Função Interativa: A relação entre a imagem e o espectador

A função interativa diz respeito à maneira como as imagens estabelecem relações com os observadores. Ela abrange aspectos como distância social, ângulo de visão, contato visual e direção do olhar. Esses elementos são importantes para definir o tipo de envolvimento que a imagem propõe ao espectador. Kress e van Leeuwen (2006) identificam três níveis de distância social: (a) Íntima (close-up ou muito próximo ao rosto ou detalhe do corpo); (b) Pessoal (distância moderada, como de um retrato comum); e (c) Social (planos mais abertos, que mostram o participante em contexto).

Essas escolhas afetam como o espectador se sente em relação ao conteúdo – mais próximo, envolvido, ou mais distante e neutro. O ângulo de visão também influencia a percepção: uma imagem de cima pode indicar superioridade ou fragilidade do participante; uma imagem de baixo, poder; enquanto o olhar direto (contato visual) cria envolvimento e interpelação.

Nas *thumbnails*, esses recursos são manipulados com intencionalidade persuasiva: olhares diretos, rostos em close-up e ângulos dramáticos,

como veremos nas análises, são estratégias comuns para capturar a atenção e provocar identificação emocional com o conteúdo.

6.3. Função Composicional: A organização visual do significado

A função composicional diz respeito à forma como os elementos visuais são organizados dentro do espaço gráfico para formar uma mensagem coesa e coerente. Para Kress e van Leeuwen (2006), os principais aspectos estudados na função composicional são:

- Enquadramento (*Framing*): maneira como os elementos visuais são contidos ou delimitados dentro de um layout. O enquadramento afeta a atenção do espectador e a interpretação do conteúdo visual, criando limites visuais que destacam ou isolam partes específicas do design.

- Posicionamento (*Positioning*): localização precisa dos elementos visuais dentro de um layout. O posicionamento influencia como os espectadores percebem a relação entre os diferentes componentes do design, ajudando a guiar o olhar e a criar uma estrutura lógica e organizada para a apresentação visual.

Enquadramento e posicionamento organizam valores informacionais codificados em elementos dispostos espacialmente em diferentes “zonas” de uma imagem (Figura 6).

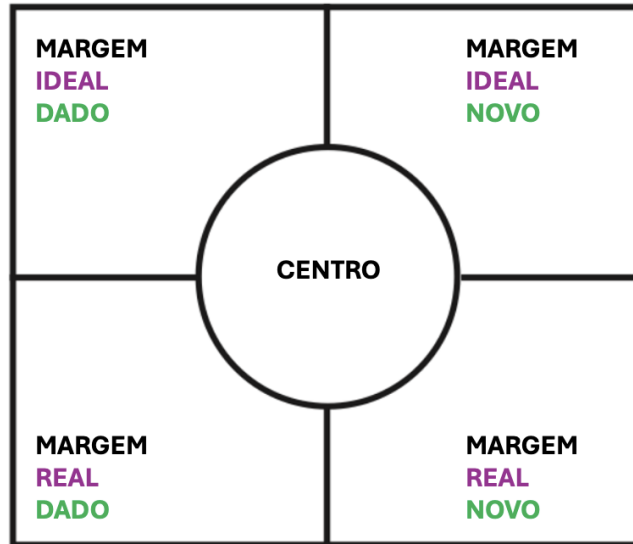


Fig.6 – Função Composicional

Fonte: Autores, com base em Kress e van Leeuwen (2006).

A Figura 6 resume algumas categorias observadas na composição da informação:

- Dado e Novo: o dado refere-se à informação que é considerada familiar, conhecida ou de fácil acesso para o espectador. Normalmente é posicionada à esquerda do layout visual. O elemento novo refere-se à informação nova, desconhecida ou importante que o espectador precisa prestar atenção. É geralmente colocada à direita do layout visual.

- Ideal e Real: a parte superior de um layout visual frequentemente apresenta informações ideais, abstratas ou teóricas (Ideal), enquanto a parte inferior mostra informações mais concretas, práticas ou reais (Real).

- Saliência: destaque dado a certos elementos visuais para captar a atenção do espectador, por exemplo, o elemento do centro da imagem. A saliência também pode ser manipulada através do tamanho, cor, contraste, posição, foco, e outros fatores visuais que tornam um elemento mais perceptível do que outros. A Saliência está muito associada com o enquadramento centro/margem, pois alguns layouts utilizam a polaridade central para destacar o elemento central como o mais importante, com elementos marginais fornecendo suporte ou contexto adicional.

Na Figura 4, as linhas, bordas e espaços em branco demonstram os espaços que podem ser usados para criar relações ou distinções entre diferentes elementos visuais, influenciando a interpretação e a coerência do layout. O enquadramento e o posicionamento criam as zonas de informação pela divisão do layout em zonas que agrupam informações relacionadas.

Esses princípios podem ser observados com clareza nas *thumbnails*, que utilizam, como descreveremos na análise, hierarquia visual para guiar o olhar do usuário, destacar elementos centrais (como rostos ou textos chamativos), e combinar cores e composições para reforçar sentidos e emoções. Por exemplo, cores quentes como vermelho e amarelo frequentemente são associadas à urgência, emoção e excitação, enquanto tons frios podem transmitir calma ou mistério (DONDIS; CAMARGO, 1997).

Diante da análise das funções representacional, interativa e composicional propostas pela GDV, torna-se evidente que as *thumbnails* operam como construções verbo-visuais complexas, estrategicamente organizadas para captar a atenção e persuadir o público. No entanto, para compreender plenamente o funcionamento desse gênero digital no ecossistema das redes, é necessário também considerar o contexto mais amplo em que ele circula. As *thumbnails* não refletem apenas escolhas estéticas ou funcionais. Elas estão imersas em práticas comunicativas próprias da cultura digital. Assim, a próxima seção recorre aos aportes da Linguística da Internet, com o objetivo de explorar como as particularidades das interações online, os modos de produção e consumo de conteúdo e os gêneros emergentes influenciam a constituição e circulação das *thumbnails* nas plataformas digitais.

7. Linguística da Internet: o estudo da linguagem nos meios digitais

Nas últimas décadas, a internet transformou os meios de comunicação e, além disso, as formas como a linguagem é produzida, distribuída e interpretada. Para Crystal (2011, p. 1),

Onde quer que encontremos linguagem, encontraremos linguistas. É para isso que os linguistas existem: para buscar, descrever e analisar as manifestações da linguagem em todos os lugares. Portanto, ao nos depararmos com o maior banco de dados linguístico que o mundo já viu, seria de se esperar que linguistas o estivessem explorando, para entender o que está acontecendo. Isso já começou a acontecer. E, como consequência, um novo campo está emergindo: a Linguística da Internet.

Com o advento das tecnologias digitais, esse campo de investigação ampliou-se para abarcar manifestações mediadas por computadores e dispositivos móveis. A Linguística da Internet (doravante LI) é, portanto, voltada ao estudo das práticas linguísticas que ocorrem nos meios eletrônicos e digitais.

A LI não substitui os ramos tradicionais da Linguística – como sintaxe, semântica, morfologia ou sociolinguística –, mas desloca o foco analítico para os contextos comunicacionais mediados por tecnologia. Desse modo, fenômenos como o multiculturalismo, a multimodalidade e a aceleração da comunicação são considerados em termos de seu impacto nos modos de significar. Como destaca Crystal (2011), os desafios enfrentados por linguistas nesse campo decorrem justamente da simultaneidade de códigos, da diversidade cultural e das novas formas de textualidade que se impõem nos ambientes digitais.

O caso das *thumbnails*, foco deste estudo, ilustra bem a necessidade dessa abordagem. Como elemento visual-textual recorrente em plataformas como o *YouTube*, as *thumbnails* exemplificam um tipo específico de *output*. Este termo (*output*) refere-se a diferentes formatos discursivos que surgem com o uso da internet – como e-mails, blogs, chats, tweets, entre outros – e que, segundo Crystal (2011, p.9-11) ainda não apresentam uma homogeneidade linguística suficiente para serem definidos, de forma segura, como gêneros, variedades, registros ou dialetos. Segundo o autor, utilizar essas classificações implicaria assumir uma coerência linguística que ainda precisa ser demonstrada por meio de análises empíricas.

Dentro destes *outputs*, as *thumbnails* se integram para condensar informações verbais e visuais estrategicamente, orientadas ao clique, à curiosidade e ao engajamento. Em ambientes regidos por algoritmos e disputas de atenção, como as redes sociais, esses elementos funcionam como porta de entrada e síntese do conteúdo audiovisual.

É nesse sentido que a Linguística da Internet contribui para a análise das *thumbnails*: ao considerar os *outputs* digitais possuem estruturas comunicativas moldadas por dinâmicas próprias da cultura digital como, por exemplo, a velocidade da atualização, a multimodalidade, o anonimato, a interatividade não obrigatória e a persistência relativa do conteúdo (CRYSTAL, 2011). Estudos de Miller e Shepherd (2004), Askehave e Nielsen (2005; 2022) Silva (2013), Page (2019), Azevedo (2020) e Vokhidova (2024), entre outros, demonstram como os textos digitais exigem que se repense a própria noção de linguagem, levando em conta fenômenos como o *displacement* digital – a capacidade da linguagem de transcender tempo e espaço através de mensagens, vídeos, áudios, gifs, memes, entre outros formatos.

Além disso, a escrita e a fala na internet se mesclam em graus variados, dependendo do tipo de plataforma e da interação envolvida. Enquanto em artigos científicos online pode-se manter uma linguagem tradicional e formal, em comentários de vídeos ou chats ao vivo observa-se uma linguagem próxima à oralidade, com marcas de urgência, emoção e espontaneidade. De acordo com Crystal (2011), no contexto digital,

É mais realista pensar na fala e na escrita como pontos finais de um continuum multidimensional, dentro do qual as variedades podem ser localizadas como sendo 'mais ou menos parecidas com a fala' e 'mais ou menos parecidas com a escrita' (CRYSTAL, 2011, p. 19 – tradução livre).

Essas diferenças tornam-se especialmente relevantes para o estudo das *thumbnails*, que muitas vezes combinam recursos verbais sintéticos e apelativos com imagens emocionalmente expressivas, refletindo as lógicas interacionais da comunicação digital.

A LI, portanto, oferece um arcabouço necessário para compreender como as *thumbnails* operam como práticas linguísticas multimodais, situadas em um ecossistema marcado pela fluidez dos gêneros, pela hibridez das linguagens e pela constante renegociação das normas comunicativas. Dentro dessa abordagem, Susan Herring (2007) oferece uma importante contribuição teórica ao propor um modelo de análise que articula facetas tecnológicas e facetas sociais da comunicação digital, reconhecendo que os textos da internet são moldados tanto pelas características técnicas das plataformas quanto pelas práticas sociocomunicativas dos usuários.

7.1. As facetas sociais e tecnológicas

Ao estudar a linguagem no mundo digital, Herring (1996, 2007) argumenta que os outputs digitais, como blogs, e-mails, fóruns, posts em redes sociais ou *thumbnails*, não podem ser analisados apenas com base em suas

semelhanças com gêneros tradicionais. A autora propõe que cada *output* deve ser compreendido em função de um conjunto de parâmetros: as facetas que configuram seu modo de funcionamento. Essas facetas funcionam como eixos contrastivos que permitem descrever como as mensagens são construídas, compartilhadas e interpretadas em diferentes ambientes digitais para a análise de gêneros emergentes.

Os Quadros 2 e 3, a seguir, resumem as facetas sociais e tecnológicas.

Faceta	Descrição
Granularidade	A natureza das unidades transmitidas (linhas, mensagens, caracteres).
Tamanho	O número de caracteres que um sistema permite em uma mensagem.
Canais	Os meios de transmissão: gráficos animados, vídeo, áudio ou uma combinação.
Sincronicidade	Comunicação síncrona, assíncrona ou híbrida.
Identidade	Participantes anônimos ou identificados.
Formato	As mensagens são acessíveis pública ou privadamente.
Adaptação	O sistema permite que o conteúdo seja filtrado, citado ou modificado.
Persistência	O período de tempo que as mensagens permanecem no sistema.

Quadro 2 – Facetas Tecnológicas

Fonte: Autores, com base em Herring (2007)

Faceta	Descrição
Estrutura de Participação	Número de participantes ativos ou potenciais, tempo e velocidade de interação, se ocorre em público ou privado, real ou fictício.
Código	Convenções linguísticas compartilhadas (abreviações, piadas internas, variações ortográficas, etc.). Cada output tem seu código.
Normas de Linguagem	Línguas ou variedades linguísticas usadas (faladas e/ou escritas).
Características dos Participantes	Fatores sociolinguísticos como idade, gênero, formação, cultura, crenças, profissão, etc.
Propósito	Motivo(s) para a mensagem.
Atividades	Meios utilizados para atingir o propósito (por exemplo, texto, imagens, vídeos, etc.).

Tópico	Tipo de conteúdo considerado apropriado para a mensagem, dependendo do output.
Tom	O espírito ou intenção da interação (formal, informal, irônico, humorístico etc.).
Normas de Organização	Como os participantes se organizam (moderação, entrada de novos membros, controle de tópicos).
Normas de Adequação Social	Regras de conduta aceitas na interação (netiqueta, filtros, proibições, etc.).

Quadro 3 – Facetas Sociais
 Fonte: Autores, com base em Herring (2007)

O modelo de Herring é especialmente produtivo para a análise de gêneros digitais, como as *thumbnails*, uma vez que consideram tais gêneros como elementos visuais condicionados pelas plataformas em que circulam (por exemplo, o *YouTube*) e pelos padrões de interação dos usuários. Diferentes *outputs* impõem diferentes combinações de facetas: os tweets, por exemplo, apresentam limitação de caracteres, são altamente públicos e tendem à interação veloz; os e-mails, por sua vez, costumam seguir normas formais e circulam em ambientes restritos. Já os vídeos em plataformas como *YouTube*, *TikTok* ou *Instagram* reúnem diversos canais comunicativos (imagem, som, texto) e permitem comentários, curtidas, compartilhamentos e respostas, operando com fortes graus de multimodalidade, interatividade, persistência variável, etc.

Nesse contexto, a *thumbnail* se destaca como um componente verbo-visual altamente estratégico. Embora pequena, essa imagem carrega informações densas que mobilizam múltiplas facetas tecnológicas (como canais visuais, persistência e adaptação gráfica) e sociais (como tom, propósito e normas de adequação). Ela atua como uma síntese visual do conteúdo do vídeo, mas também como um gatilho de engajamento, apelando para as emoções, a curiosidade e o desejo de pertencimento do público digital. A seguir, apresentamos a metodologia do estudo,

articulando os conceitos da Linguística da Internet, da Gramática do Design Visual (GDV) e da Estrutura Potencial Genérica (EPG).

8. Metodologia: dados e procedimentos de estudo

Esta pesquisa caracteriza-se como um estudo qualitativo, de natureza interpretativista, com foco na análise de gêneros digitais multimodais em contextos de alta circulação online. Por tratar-se de uma investigação baseada exclusivamente em dados públicos e acessíveis ao usuário comum na internet (*YouTube*), não houve necessidade de submissão ao Comitê de Ética em Pesquisa, conforme resolução vigente do Conselho Nacional de Saúde.

O corpus é composto por *thumbnails* e respectivos títulos de vídeos obtidos a partir da seção “Em alta” da plataforma *YouTube*, durante os meses de maio de 2025, conforme surgiam vídeos “em alta” nas diferentes categorias estabelecidas. As *thumbnails* extraídas dos canais de *podcast* tiveram uma pequena diferença, pois foram selecionados os últimos vídeos (que eram, especificamente, *podcasts*) postados no dia 01/06/2025 dos maiores canais deste nicho em relação à quantidade de inscritos. Os vídeos foram organizados em categorias temáticas amplas com base nos nichos de maior incidência na plataforma (Quadro 4):

Categoria	Quantidade	Canal/ Criador
Entretenimento	5	Lucas Tyty, Gabriel e Shirley, Banheiristas, Julio Cocielo, Enaldinho
Games	5	Geleia, Jazzghost, vinicius13, Mendrake, Problems
Reação/Comentário	5	Fábio Sormani, Fala Cap, BRKsEDU, Funky Black Cat, Regis Tadeu
Música e Arte	5	MC IG, Grupo Menos é Mais, VEIGH, Mnet K-POP, Meno Tody

Vlog	5	NiinaSecrets, Lorryne Oliveira, Franksângela, Kathy Castricini, Lactea
Jornalismo	5	Gameplayrj, POD DELAS, Eu Joguei - Cross, Celso Russomanno, UOL Esporte
Esporte	5	Free Fire Esports Brasil, VALORANT Esports BR, Kings League Brazil, JogaNoGrau, CazéTV
Podcast	5	POD DELAS, Mais que 8 minutos, Inteligência LTDA, Flow Podcast, PODPAH
Educação/Curiosidades	5	Aero Por Trás da Aviação, Tiago Jácomo, TerraVerde Português, Erico Rocha, No paddock da F1 com a Ju

Quadro 4 – Corpus de análise

Fonte: Autores

A seleção dos vídeos priorizou o próprio *ranking* disponibilizado no “em alta”, sendo as categorias separadas pelos vídeos em si, não levando somente em consideração o estilo/categoria que o criador normalmente produz seu conteúdo ou se consolidou.

A coleta foi realizada por meio de capturas de tela (*screenshots*), organizadas em banco de dados com as seguintes informações: nome do canal, link do vídeo, categoria temática, data de postagem e imagem da thumbnail.

A análise do corpus foi dividida em dois eixos complementares, conforme os referenciais adotados:

- 1) Análise do texto escrito e visual (títulos + *thumbnails*): com base na Linguística Sistêmico-Funcional (Halliday, 1994), com foco nas três metafunções (ideacional, interpessoal e textual), a fim de compreender como os significados são organizados para representar experiências, posicionar interlocutores e estruturar a mensagem. A análise do texto visual foi fundamentada na Gramática do Design Visual (Kress; van Leeuwen, 2006), com ênfase nas funções representacional, interativa e composicional. Foram observados aspectos como

- enquadramento, uso de cores, vetores, expressões faciais, layout e saliência. Salientamos a necessidade de análise conjugada de título e *thumbnails* pois ambos atuam na construção da persuasão e composição multimodal do gênero.
- 2) Análise do gênero: orientada pelos conceitos de Estrutura Potencial Genérica (Hasan, 1989) e pelas facetas sociais e tecnológicas propostas por Herring (2007), com o intuito de identificar padrões recorrentes nas construções verbo-visuais do gênero thumbnail em contextos de alta concorrência comunicativa.

As categorias de análise foram organizadas em planilhas, permitindo a sistematização dos dados, a observação de regularidades, contrastes e tendências, e a integração entre os diferentes aspectos composicionais, discursivos e contextuais das *thumbnails* e títulos analisados.

4. Análise dos Dados

Em termos gerais, as análises demonstraram que a multimodalidade é a marca mais preponderante nas *thumbnails* e sua integração com os títulos.

4.1 . Análise do texto escrito e visual

A representação da experiência teve a proporção descrita no Gráfico 1:

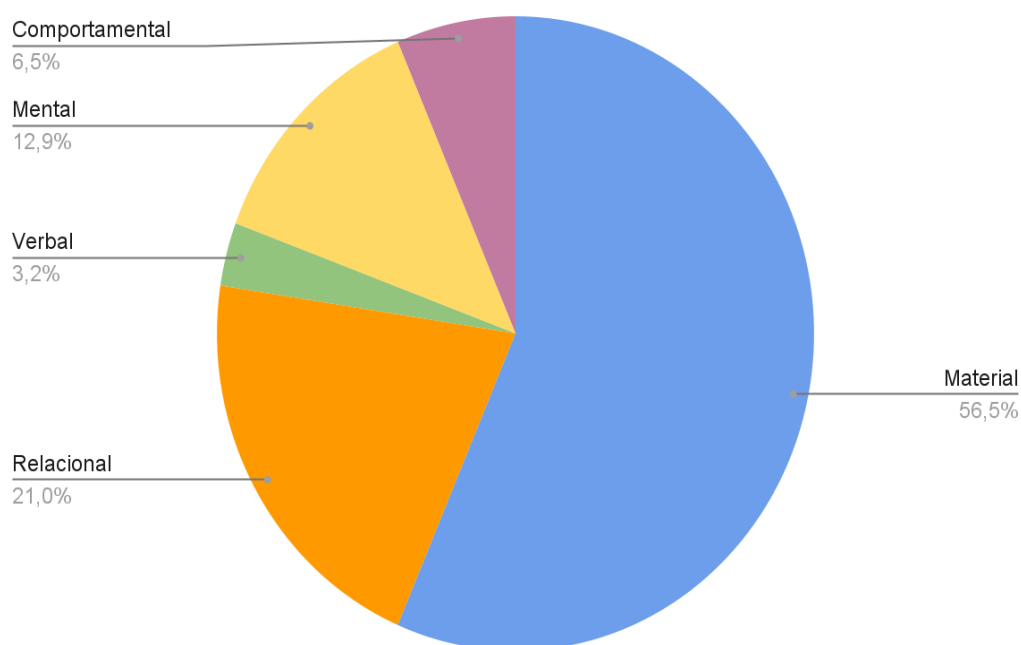


Gráfico 1 – Incidência de Processos no Corpus

Fonte: Autores

O gráfico mostra que, no corpus analisado, há predominância de processos materiais (56,5%), representando ações concretas, e processos relacionais (21%), de atribuição ou identificação, seguidos de proporções menores de mentais (12,9%), comportamentais (6,5%) e verbais (3,2%). A forma como representamos a experiência participa decisivamente nas funções discursivas do gênero.

Ao lado de significados interpessoais, avaliativos e atitudinais, processos, participantes e circunstâncias também participam nas práticas de *online bait*. A prática de *bait* (mais conhecida no meio digital como *clickbait*) é uma estratégia discursiva e multimodal usada para induzir o usuário a clicar em um conteúdo por meio de elementos que despertam curiosidade, emoção ou senso de urgência, muitas vezes apresentando a informação de forma parcial, sugestiva ou exagerada (MORALES; MARINO-JIMÉNEZ, 2024).

No contexto de *thumbnails* e títulos do YouTube, isso envolve algumas características principais: (a) o apelo à curiosidade (*curiosity gap*) implica em mostrar o suficiente para gerar interesse, mas reter parte da informação para forçar o clique; (b) a linguagem superlativa se realiza pelo uso de adjetivos e advérbios intensificadores; (c) a promessa de revelação ou exclusividade implica em transmitir que o vídeo oferece algo raro, secreto ou inédito; e (d) a estrutura de suspense é criada ao deixar perguntas sem resposta no enunciado.

A Figura 7 reúne os principais *bait*s (*clickbaits*) identificados no corpus:



Fig. 7 – Tipos recorrentes de *clickbaits* nos títulos das *thumbnails*
Fonte: Autores

A Figura 7 evidencia que a integração entre elementos verbais e visuais é fundamental para alinhar texto e imagem na criação de um mesmo gatilho persuasivo. Nessa prática, mobiliza-se também a metafunção ideacional. O Quadro 5, a seguir, apresenta como os processos materiais – prioritariamente – e relacionais (estes, muitas vezes realizados sem cópula verbal, por justaposição de participantes e recuperação contextual entre enunciado e imagem) desempenham função discursiva e papel central na performatividade do *bait*.

Processo	Descrição	Exemplos	Função Discursiva	Papel no Bait
Material	Comunica ação imediata, movimento, mudança ou conquista, criando senso de dinamismo e urgência.	“Completei a coleção...”, “Testando os piores produtos...”, “Peguei o prêmio...”	Induz expectativa de ver a ação e seu resultado, incentivando o clique.	Promete “o que vai acontecer/aconteceu” (ação).
Relacional	Estabelece quem é/foi algo, ou qual é a qualidade/atributo de alguém ou algo.	“Lady Gaga - A Maior Farsa Pop do Ano?”, “Vou ser vovó...”	Reforça identidade, status ou relevância, posicionando o tema como valioso.	Posiciona “quem/qual é o objeto ou pessoa” (atribuição).
Integração Material + Relacional	Combina ação com atribuição, criando narrativa completa, geralmente por meio de processos implícitos	“Novo golpe em posto de gasolina”, “Quem é o jogador profissional? feat. Leo Moura”	Apresenta evento e valor social/emocional simultaneamente	Cria um <i>gancho duplo</i> que une expectativa de ação ao reconhecimento de relevância.

Quadro 5 – Processos preponderantes e função discursiva na construção do Bait

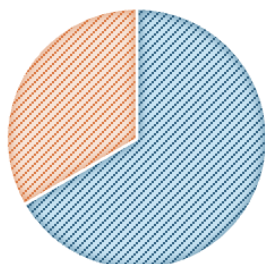
Fonte: Autores

Nas *thumbnails* analisadas, estabelece-se uma relação típica de gêneros multimodais: o título cumpre a função de apresentar a promessa verbal, geralmente mobilizando processos materiais e relacionais e recorrendo a gatilhos lexicais específicos, enquanto a *thumbnail* reforça visualmente o apelo por meio de rostos, expressões, objetos-chave e uso estratégico de cores contrastantes. Em conjunto, esses elementos configuram um gênero híbrido que integra recursos próprios da manchete jornalística, da publicidade e do design gráfico.

A análise da metafunção interpessoal do enunciado escrito foi baseada nos critérios de modalidade e Avaliatividade (MARTIN; WHITE, 2005).

MODALIDADE

■ Implícita ■ Explícita



AVALIATIVIDADE

■ Julgamento Positivo Implícito
 ■ Julgamento Positivo Explícito
 ■ Julgamento Negativo Explícito
 ■ Julgamento Negativo Implícito
 ■ Afeto Positivo Explícito

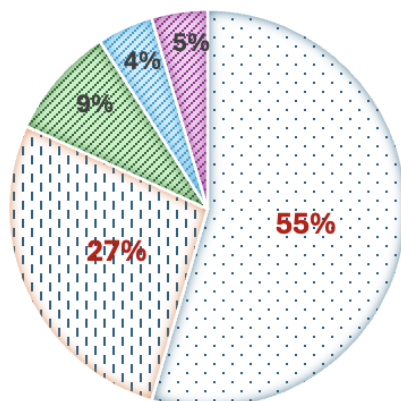


Gráfico 1 – Modalidade e Avaliatividade nos *thumbnails*
 Fonte: Autores

A análise da Modalidade revela uma predominância da forma implícita (66,7%) em relação à explícita (33,3%), indicando que, na maior parte das ocorrências, a intenção persuasiva não é declarada de maneira direta, mas inferida a partir do contexto, das escolhas lexicais e da relação entre título e imagem. Esse recurso confere maior sutileza ao discurso e convida o espectador a preencher lacunas interpretativas, estimulando a curiosidade. Por exemplo, títulos como “*Torrei meu dinheiro em...*” ou “*Quem é o jogador profissional?*” não explicitam toda a informação, mas sugerem uma ação ou revelação. Este recurso enquadra-se numa estratégia de *bait* baseada no efeito de antecipação.

Quanto à Avaliatividade, observa-se que Julgamento Positivo Implícito é a categoria mais recorrente (55%), seguido por Julgamento Positivo Explícito (27%), enquanto avaliações negativas, como Julgamento Negativo Explícito (9%) e Julgamento Negativo Implícito (4%). As manifestações de Afeto e Apreciação são menos frequentes. A predominância de julgamentos positivos, especialmente implícitos, indica que os títulos e *thumbnails* tendem a reforçar qualidades, conquistas ou

relevância dos participantes humanos – ao menos nas categorias/nichos analisados –, muitas vezes sem uso direto de adjetivos avaliativos, mas por meio da seleção de eventos e enquadramentos visuais. Assim, enunciados como “Área restrita do aeroporto com...” ou “Elon Musk X Trump” combinam a construção de autoridade e exclusividade com uma valorização indireta. Funcionam pois como gatilhos persuasivos para atrair o clique.

A análise da GDV trouxe à luz diversos aspectos de padrões e variações associados às *thumbnails*. As principais delas vêm resumidas a seguir, nas Figuras 8,9 e 10, enquanto o Quadro 6 as complementa.



Fig. 8 – Saliência nas *thumbnails*

Fonte: Autores

3) Framing (enquadramento e segmentação)

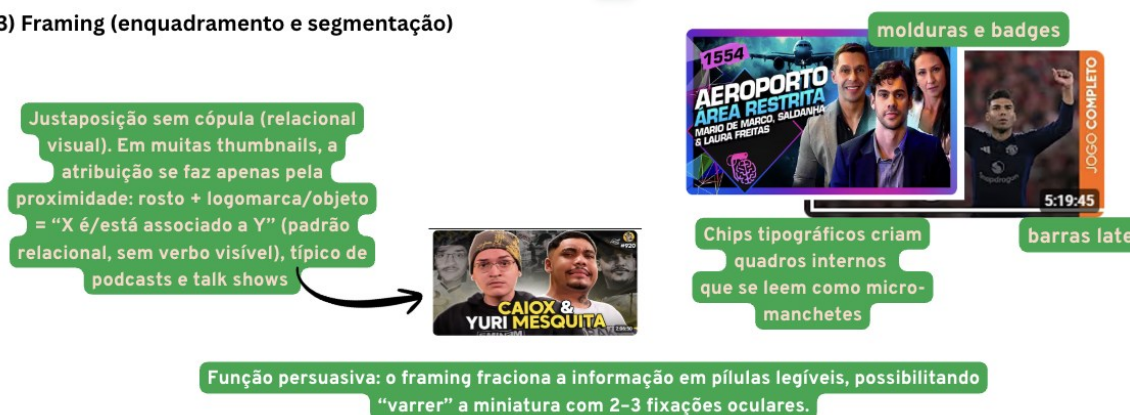


Fig. 9 – Framing nas *thumbnails*
Fonte: Autores

1) Valor de informação (posição)



Fig. 10 – Valor da Informação nas *thumbnails*
Fonte: Autores

Tópico (GDV)	Ponto-chave	Descrição operacional	Função persuasiva
Centro–margem	Centro dominante	Participantes em <i>close</i> no centro; margens para logo/ícones/cenário.	Fixa “quem fala”; foco imediato no rosto.
Hierarquia tipográfica	Headline -> subtítulo -> selos	1) Título pesado; 2) gancho curto; 3) #episódio/“ao vivo”.	Governa leitura macro -> micro; alta legibilidade.
Relações intermodais	Material no título; prova na imagem	Verbo de ação no título; evidência visual (gesto/objeto) na <i>thumb</i> .	Divide tarefas; reforça promessa + prova.
Relações intermodais	Relacional por justaposição	Rosto + logotipo/objeto = “X é Y” sem cópula verbal.	Identificação rápida; autoridade.
Relações intermodais	Curiosity gap espacial	Ideal (fundo/topo) sugere; Real (faixa inferior) confirma “quem/tema”.	Mantém lacuna informativa; incentiva clique.
Arquétipos	Duel/Face-off	Split cromático; dois perfis; “X” entre eles; <i>badges</i> .	Dramatiza conflito → <i>bait</i> de status/polêmica.
Arquétipos	Host + Guest(s) com cartuchos	Host à esquerda (Dado); convidados à direita (Novo); cartuchos de tema/#.	Autoridade + novidade -> exclusividade/credibilidade.
Arquétipos	Action + Object-ícone	Gesto do participante; objeto ampliado; emojis/setas como vetores.	Resultado de ação -> <i>bait</i> de recompensa.
Gaze & distância	Olhar para a câmera / close-up	Endereço direto; distância íntima (planos fechados).	Empatia e legibilidade em telas pequenas.
Vetores	Setas/gestos direcionais	Gestual e ícones conduzem ao texto-gatilho/objeto.	Orienta a varredura ocular.
Cromia	Duotons complementares	Para contraste e oposição.	Codifica tensão/energia; separa lados.
Branding	Cores-assinatura + selos	Paleta do canal + cartuchos consistentes.	Reconhecimento instantâneo no feed.
Síntese funcional	Localiza	Dado/Novo + centro identificam “quem/quê/onde”.	Reduz custo de decisão.
Síntese funcional	Promete/Mostra	Ideal promete; Real confirma “quem/tema”.	Mantém curiosidade sem spoiler.
Síntese funcional	Focaliza gatilho	Saliência do rosto + palavra-chave.	Maximiza CTR.
Síntese funcional	Legítima	Framing + selos de live/#.	Sinaliza valor/autoridade.

Quadro 6 – Aspectos da GDV nas *thumbnails*

Fonte: Autores

4.2 . Análise do gênero

Esta parte da análise foi orientada (i) pela Estrutura Potencial Genérica – EPG (Hasan, 1989), para descrever estágios/movidas obrigatórias e opcionais do gênero *thumbnail*; e (ii) pelas facetas sociais e tecnológicas de Herring (2007), para explicar como posições sociais, estilos de fala e afordâncias de plataforma condicionam essas escolhas verbo-visuais em contexto de alta concorrência comunicativa (*feeds* saturados e decisão de clique em milissegundos).

EPG do gênero *thumbnail* – no corpus, a macro-organização recorre a um conjunto relativamente estável de movidas:

Elemento	Categoria (EPG)	Descrição/ função
(M1) <i>Identificação visual do participante</i>	<i>Quase obrigatória</i>	elemento/rosto(s) em <i>close</i> como âncora central da leitura.
(M2) <i>Headline tipográfica</i>	<i>Obrigatória</i>	unidade verbal curta e de alto peso (tema/nome/resultado) que funciona como gatilho.
(M3) <i>Sinalização de marca/série</i>	<i>Opcional</i>	logotipo do canal e/ou cores-assinatura para reconhecimento rápido.
(M4) <i>Evidência visual do tópico</i>	<i>Opcional – recorrente</i>	objeto-ícone, cenário ou gesto que “prova” a promessa do texto.
(M5) <i>Metadados de circulação</i>	<i>Opcional</i>	elementos como #episódio/“ao vivo”/placar.

Quadro 7 - Estrutura Potencial Genérica das *thumbnails*
Fonte: Autores

No aspecto sequenciais, a ordem típica de elementos, com grande variabilidade, é [M1+M2] -> M3 -> M4 -> M5/M6, com M1 e M2 como núcleo não-elidível. Em termos intermodais, processos materiais tendem a ser veiculados em M2 (título/headline), enquanto atribuições relacionais emergem, geralmente por justaposição visual em M1/M4.

Ao investigar as *thumbnails* quanto às facetas sociais, predomina código informal/coloquial (brevidade, superlativos, perguntas retóricas) e avaliatividade positiva (frequentemente implícita), associada à prova social (convidados de prestígio, números de episódio, selos “live”). A relação interpessoal é, geralmente, simulada via olhar (*gaze*) direto e distância íntima (*close-ups*), que aproximam audiência e criador; a posição de autoridade é construída por *branding* consistente e por *frames* que legitimam (barras, cartuchos, selos) e são rotineiramente aplicados como identidade visual.

As Facetas tecnológicas também apresentaram resultados relevantes. As afordâncias do *YouTube/feeds* móveis impõem formato 16:9, sobreposição do tempo no canto inferior direito e miniaturização em telas pequenas. Disso derivam: (a) centralização do pico de saliência (e.g., rosto + palavra-gatilho), (b) hierarquia tipográfica, comumente explicativa ou com função de *bait* (1–4 palavras de alto contraste), (c) duotons e hierarquização cromática complementares para segmentar lados/temas e (d) flexibilidade no valor de informação: embora o par Dado/Novo siga majoritariamente a ordem canônica (esq.=Dado/dir.=Novo), há inversões estratégicas quando a novidade precisa ocupar o ponto de entrada do olhar; no eixo Ideal/Real, a horizontalidade favorece a faixa central saliente e, quando explicitado, a tendência é manter a ordem canônica (Ideal acima/Real abaixo), com inversões pontuais por legibilidade e colisão com metadados.

Ao final da análise do corpus, dentro desse contexto, foi possível identificar alguns padrões recorrentes:

- Centro dominante com participantes;
- headline pesada que muitas vezes substitui a cópula (relações “X = Y” realizadas graficamente);

- divisão de trabalho texto X imagem: promessa no verbal, evidência no visual;
- três arquétipos composicionais: *Face-off* (oposição cromática/duas faces), *Host + Guest(s)* com aspectos (Dado-Novo, em posicionamento variável) e *Action + Object-ícone* (gesto, objeto ampliado, vetores);
- olhar direto, *close-up* e vetores (setas/gestos) guiando a varredura ao gatilho. Processos transacionais são os mais comuns, menos em contexto de *bait*, quando os reacionais são preponderantes;
- *branding* consistente (paleta/logos/selos) garantindo reconhecimento seriado.

A integração da EPG com as facetas de Herring (2007) mostra que a *thumbnail* opera como gênero híbrido e estrategicamente economizado. A implicitude do gênero favorece lacunas informacionais com efeito de *bait*s (de vários tipos). Nossas análises demonstraram a tendência de uma configuração discursiva e funcional de: localiza (centro + Dado/Novo), promete (Ideal) e mostra (Real) sem resolver o enigma, focalizando um único gatilho saliente (e.g., rosto + palavra-chave) e legitimando (opcionalmente) com *framing* e marca.

Essa gramática composicional responde simultaneamente às pressões sociais (disputa por atenção e construção de *ethos*) e às restrições técnicas (miniatura, algoritmo, legibilidade). Tais padrões consolidam a alta eficiência persuasiva das *thumbnails* em ambientes de concorrência comunicativa intensa.

5. Algumas Considerações

Este estudo buscou compreender as *thumbnails* e títulos do *YouTube* como gêneros multimodais. Nesse sentido, articulamos as características visuais

e linguísticas que desempenham papel na persuasão e no engajamento digital do gênero.

A análise evidenciou padrões composicionais recorrentes, como a centralidade de imagens, o uso de contrastes cromáticos e a associação com títulos que exploram estruturas sintáticas simples e impactantes. Ao menos no contexto analisado – canais brasileiros de categorias bem segmentadas e de grande profusão (em alta) –, as estratégias empregadas pelos criadores/autores apresentaram uma organização estética e persuasiva bem estabelecida (em elementos obrigatórios e opcionais) com a finalidade de maximizar a visibilidade e ampliar o potencial de cliques e compartilhamentos dos vídeos.

Os resultados também mostraram que tais gêneros reproduzem e ressignificam práticas comunicativas tradicionais, como manchetes jornalísticas e peças publicitárias, ao integrarem novas tecnologias de edição e recursos multimodais. Além disso, verificou-se a recorrência de apelos emocionais, promessas de novidade e exclusividade, elementos que reforçam estratégias de *bait* e consolidam a função persuasiva das *thumbnails* no ecossistema digital.

Assim, conclui-se que as *thumbnails* constituem, na atual conjuntura, instrumentos discursivos extremamente relevantes e capazes de moldar a experiência de consumo de conteúdo e influenciar práticas culturais na sociedade em rede.

Esperamos que este trabalho contribua para ampliar as discussões sobre gêneros digitais e multimodais, muitas vezes desprivilegiados como as *thumbnails*. Acreditamos que pesquisas aprofundarão o entendimento sobre o gênero, além de seus impactos sociais, ideológicos e comunicativos.

REFERÊNCIAS

ASKEHAVE, Inger; NIELSEN, Anne Ellerup. Digital genres: A challenge to traditional genre theory. **Information Technology & People**, v. 18, n. 2, p. 120–141, 2005.

ASKEHAVE, Inger; NIELSEN, Anne Ellerup. Digital genres: What they are, what they do, and why we need to better understand them. **Journal of Pragmatics**, v. 198, p. 19–30, 2022.

CECI, Laura. **O papel do YouTube na economia da atenção e no marketing digital**. São Paulo: Editora Mídia e Sociedade, 2025.

CRYSTAL, David. **Internet Linguistics: A Student Guide**. London: Routledge, 2011.

DONDIS, Donis A.; CAMARGO, Jefferson Luiz. **Sintaxe da linguagem visual**. São Paulo: Martins Fontes, 1997.

GODIN, Seth. **All marketers are liars: The power of telling authentic stories in a low-trust world**. New York: Portfolio, 2005.

HALLIDAY, M. A. K. **An Introduction to Functional Grammar**. 2. ed. London: Edward Arnold, 1994.

HALLIDAY, M. A. K.; HASAN, Ruqaiya. **Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective**. Oxford: Oxford University Press, 1989.

HALLIDAY, M. A. K.; MATTHIESSEN, Christian M. I. M. **Halliday's Introduction to Functional Grammar**. 4. ed. London: Routledge, 2014.

HASAN, Ruqaiya. The structure of a text. In: HALLIDAY, M. A. K.; HASAN, R. **Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective**. Oxford: Oxford University Press, 1989. p. 52–69.

HERRING, Susan. **Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives**. Amsterdam: John Benjamins, 1996.

HERRING, Susan C. A faceted classification scheme for computer-mediated discourse. **Language@Internet**, v. 4, 2007. Disponível em: <http://www.languageatinternet.org/articles/2007/761>. Acesso em: 05 maio 2025.

HOEFNAGEL, Joris. **Mira calligraphiae monumenta**. ca. 1570. Manuscrito iluminado produzido em colaboração com Georg Bocskay. The Getty Museum Collection. Disponível em: <https://d2aohiyo3d3idm.cloudfront.net/publications/virtuallibrary/089236212X.pdf>.

ISOLA, Maria. Gêneros digitais: práticas sociais, leitura e letramento. In: ROJO, Roxane (Org.). **Multiletramentos na escola**. São Paulo: Parábola Editorial, 2012. p. 157-181.

JANG, Ha Eun et al. The effects of thumbnail image design characteristics on user response. **Multimedia Tools and Applications**, v. 82, p. 8169–8191, 2023. DOI: <https://doi.org/10.1007/s11042-022-13566-4>.

KARWOSKI, Aline; GAYDECZKA, Beatriz. Produção textual na era digital: o blog como ferramenta de ensino e aprendizagem. In: **Anais do X Congresso de Leitura do Brasil – COLE**, Campinas, SP, 2013. Disponível em: https://www.pucsp.br/pos/letras/Revista_Letras/Links/COLE10/KARWOSKI_GAYDECZKA.pdf. Acesso em: 06 abr. 2025.

KRESS, Gunther; VAN LEEUWEN, Theo. **Reading Images: The Grammar of Visual Design**. 2. ed. London: Routledge, 2006.

MARCUSCHI, Luiz Antônio. **Gêneros textuais: definição e funcionalidade**. São Paulo: Cortez, 2002.

MARCUSCHI, Luiz Antônio. **Gêneros textuais: tipologia e ensino**. São Paulo: Parábola Editorial, 2008.

MARIANO-JIMENÉS, Mauro; MORALES, Daniel. Clickbait: Research, challenges and opportunities – A systematic literature review. **Online Journal of Communication and Media Technologies**, Facultad de Comunicación, Universidad San Ignacio de Loyola, Lima, Peru, 2024.

MARTIN, J. R. Analysing Genre: functional parameters. In: CHRISTIE, F.; MARTIN, J. R. (Ed.). **Genres and Institutions: Social Processes in the Workplace and School**. London: Cassell, 1997. p. 3–39.

MARTIN, J. R.; WHITE, P. R. R. **The Language of Evaluation: Appraisal in English**. New York: Palgrave Macmillan, 2005.

MILLER, Carolyn R.; SHEPHERD, Dawn. Blogging as social action: A genre analysis of the weblog. In: GURAK, Laura J. et al. (Ed.). **Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs**. Minneapolis: University of Minnesota, 2004.

SIMÕES, Débora Cristina. **Estrutura genérica de abstracts em português e em inglês: uma abordagem sistêmico-funcional**. São Paulo: Blucher, 2021.

TURNER, Nancy. **The Art of the Book in the Middle Ages**. Los Angeles: The J. Paul Getty Museum, 2001.

VOKHIDOVA, Tamanno. Language Features of Internet Linguistics in Modern Conditions. **European Scholar Journal**, v. 5, n. 2, p. 7–10, fev. 2024.

NOTAS DE AUTORIA

Ulisses Tadeu Vaz de Oliveira (ulisvaz@gmail.com): Ulisses Oliveira pós-doutorado no Instituto de Estudos Medievais (IEM) - Universidade Nova de Lisboa. Doutor em Linguística Aplicada e Estudos da Linguagem pela PUC-São Paulo. Trabalha na UFMS (Brasil) ensinando inglês e linguística (graduação e pós-graduação). Pesquisa: (a) TIC (aprendizado misto e em tandem); (b) Linguística Sistêmico-Funcional - estudos integrados de literatura e linguística (transitividade, persuasão e ideologia); e (c) Filologia Portuguesa e literatura comparada.

Geovana Pin Serion (geovana.serion@gmail.com): Tem experiência na área de Letras, com ênfase em Línguas Estrangeiras Modernas. Participou do evento Integra UFMS de 2023 apresentando seu trabalho de pesquisa do programa PIBIC intitulado "O SUBESQUEMA [V_QUE] NA REDE CONSTRUCIONAL DOS CONECTORES CONDICIONAIS NOS SÉCULOS XVI E XVII". Foi bolsista PIBIC pela FUNDECT orientada pelo professor Ulisses Vaz de Oliveira com o projeto de pesquisa "ANÁLISE LINGUÍSTICA E IMPACTOS SOCIAIS DOS CLICKBAITS: UMA INVESTIGAÇÃO SOBRE ESTRATÉGIAS DE ENGAJAMENTO DIGITAL", e está no último semestre do curso de Letras Português/Inglês na UFMS.

Como citar este artigo de acordo com as normas da revista?

OLIVEIRA, Ulisses Tadeu Vaz de; SERION, Geovana Pin. Compreendendo thumbnails: análise discursiva e visual no youtube. *Texto Digital*, Florianópolis, v. 21, n. 2, p. 262-305, 2025.

Contribuição de autoria

Não se aplica.

Financiamento

Não se aplica.

Consentimento de uso de imagem

Não se aplica.

Aprovação de comitê de ética em pesquisa

Não se aplica.

Licença de uso

Este artigo está licenciado sob a Licença Creative Commons CC-BY. Com essa licença você pode compartilhar, adaptar, criar para qualquer fim, desde que atribua a autoria da obra.

Histórico

Recebido em: 2 de setembro, 2025

Aprovado em: 14 de outubro, 2025