

De namorido a família: como a mistura de palavras faz o léxico se expandir

Rafael Minussi – Unifesp

IMAGINE VOCÊ CHEGANDO EM CASA
E SEU AMIGÃO PARTINDO PARA
CIMA: LATE, LAMBE E O AGARRA
COM OITO TENTÁCULOS!

E VOCÊ:

– CALMA, CALMA,

CACHOLVO!



Fonte: BUENO, Renata; CARVALHO, Beatriz. **Misturichos**. São Paulo: WMF Martins Fontes, 2012.

Resumo: A criatividade linguística manifesta-se na capacidade dos falantes de combinar elementos já disponíveis em seu léxico para produzir novas palavras. Entre os processos que revelam essa capacidade, destaca-se o *blending*, no qual duas ou mais palavras se combinam por justaposição ou sobreposição de segmentos, ou partes de palavra, como em *namorido* e *almojanta*. Este texto apresenta pesquisas sobre a formação, o processamento e o acesso lexical de *blends* no português brasileiro e europeu, desenvolvidas no âmbito do projeto B&C-Proalex. A análise de um *corpus* com mais de 500 *blends* revelou tendências estruturais, semânticas e prosódicas, indicando que a sua formação não é aleatória. Experimentos de associação de palavras e decisão lexical mostraram, ainda, que os falantes recuperam bases não necessariamente visíveis nas formações e que características prosódicas podem favorecer essa recuperação. Por fim, discute-se o estatuto dos *blends* e sua pouca exploração em materiais didáticos, defendendo sua inclusão no ensino de morfologia e no desenvolvimento da consciência morfológica na Educação Básica.

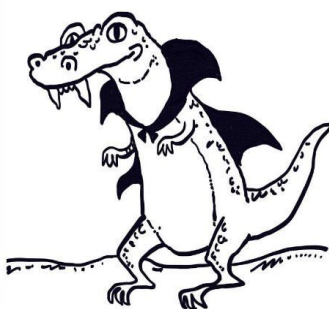
A criatividade linguística é uma das características mais importantes da linguagem humana e ela nos acompanha desde o nascimento. Podemos perceber isso em crianças que estão adquirindo uma língua, quando, por exemplo, elas começam a pronunciar frases e palavras que nunca ouviram antes e criar palavras novas, a partir das combinações de raízes, outras palavras e morfemas que estão presentes no seu léxico já adquirido. Alguns exemplos desse tipo de criatividade são: *planteiro*, *borrachar*, *atrapalhante*, *amigosa* etc. O perfil do Instagram “Frases de criança” é um prato cheio para morfólogos, com exemplos curiosos da criatividade linguística e do que podemos chamar de consciência linguística de modo geral, ou, nos casos específicos, consciência morfológica, uma vez que a criança utilizou o sistema linguístico do português, raízes e morfemas, para criar palavras.

No entanto, algo parece acontecer com as crianças quando entram na idade escolar. Algumas começam a deixar de lado essa “capacidade” criativa, e temo que a própria escola (que na realidade brasileira, por meio de um ensino “engessado”, privilegia as classificações, cobra nomenclatura e não abre espaço para reflexões linguísticas) vai, aos poucos, limitando a criatividade das crianças e jovens. Além disso, as aulas de português, infelizmente, não têm se ocupado do ensino do sistema linguístico, utilizando o conhecimento que os próprios alunos já têm sobre a língua, uma vez que os alunos já chegam na escola falando uma língua — na maioria dos casos, o português.

A observação da língua nos ajuda a entender que, ainda que a escola se esforce muito para padronizar o ensino de português, a criatividade é uma capacidade inata e, portanto, sobrevive aos anos escolares, e a morfologia parece ser um exemplo de lugar onde tal criatividade pode ser encontrada e é bastante produtiva. Há um processo de formação de palavras que tem ganhado cada vez mais espaço entre os processos produtivos de formação de palavras. Trata-se do *blending*, que origina o que chamamos de *blend* e que também recebe os nomes de cruzamento vocabular, *portmanteau*, amálgama, entre outros.

Podemos entender que o *blend* é formado a partir de duas ou mais palavras disponíveis na língua, em que há justaposição ou sobreposição de segmentos de palavras, ou uma palavra e um segmento de palavra, como ocorre em: *namorido*, *roubartilhar*, *abreijo*, *delegato*, *almojanta* e tantos outros (cf. Villalva e Minussi, 2022). Em português, encontramos alguns padrões de formação de *blends* que nos ajudam a entender essas formações e que apontam que não se trata de um processo marginal e aleatório de formação de palavras. Neste texto, vou mostrar algumas pesquisas que já foram realizadas por mim e por alguns colegas pesquisadores sobre esse processo de formação de palavras e que são reveladores do conhecimento que temos sobre as palavras e sobre como esse processo auxilia na ampliação no léxico de nossa língua.

Uma atividade interessante, que fiz diversas vezes em palestras, com os alunos da graduação e da pós-graduação, é mostrar a seguinte imagem e pedir que a nomeiem:



Fonte: Web, sem autoria.

Raramente, os interlocutores dão como resposta um nome composto como *jacaré-vampiro*. *Blends* como *dracaré* <drácula + jacaré>, *vampiré* <vampiro + jacaré>, *jacapiro* <jacaré + vampiro>, *jacacula* <jacaré + drácula>, *crocopiro* <crocodilo + vampiro>, *vampigarto* <vampiro + lagarto>, *batgarto* <batman + lagarto>, entre outros,

costumam aparecer como maioria esmagadora das respostas, sugerindo que o *blending*, pelo menos entre os adultos, é bastante produtivo e que talvez seja mais produtivo do que a composição. Esses dados nos sugerem que há diferentes padrões para a formação dessas palavras, de modo que as mesmas bases podem formar diferentes combinações como, por exemplo, vampiro e jacaré, que resultam em “vampiré” ou “jacapiro”. A pergunta que podemos fazer é: será que há alguma preferência ou algum padrão para formar *blends* em português?

Em uma pesquisa desenvolvida no âmbito do projeto B&C Proalex, envolvendo pesquisadores e laboratórios da Unifesp (LabLinC), da UFPB (LAPROL) e da Universidade de Lisboa (The Word Lab), os pesquisadores, a partir de um *corpus* de mais de 500 *blends*, coletados em redes sociais, literatura, propagandas, músicas etc., observaram a existência e certas tendências. A estrutura do *blends* foi descrita usando dois elementos: o *clip*, entendido como um segmento, ou parte de uma palavra, sem estatuto morfológico; e a palavra, quando essa está totalmente visível no *blend*. Desse modo, encontramos quatro tipos de estrutura: CP (*clip*-palavra) *delegato*, PP (palavra-palavra) *bolovo*, PC (palavra-*clip*) *cãominhada* ou CC (*clip-clip*): *seguimor*. Entre essas estruturas, a mais frequente no *corpus* foi a estrutura CP (30,7%), seguida da PP (com 29,7%). A pesquisa também revelou que a relação de coordenação (por exemplo, *namorido* <namorado + marido>), que é quando ambas as bases do *blend* podem funcionar como núcleo semântico, é a mais encontrada com cerca de 42,8%, seguida pela relação modificador-núcleo (e.g., *sofressor* <sofredor + professor>) com 40,4%.

O projeto também observou o número de sílabas das palavras formadas. Foram encontrados quatro tipos de padrões: (i) *blends* com o mesmo número de sílabas da primeira base (e.g., *advogato* <advogado + gato>; (ii) *blends* com o mesmo número de sílabas da segunda base (e.g., *pãodemia* <pão + pandemia>; (iii) *blends* que têm as bases e o *blend* com o mesmo número de sílabas (e.g., *franglês* <francês + inglês> e, por fim, (iv) *blends* com o número de sílabas diferente das bases (e.g., *livemício* (live + comício). A análise do *corpus* mostrou que 41,2% dos *blends* têm número de sílabas diferente do número de sílabas das bases dos *blends* e um equilíbrio entre *blends* com o mesmo número de sílabas da primeira base (28,4%) e *blends* com o mesmo número de sílabas da segunda base (28,2%).

A partir dos resultados do *corpus*, podemos dizer que fatores semânticos (relação núcleo-modificador), estruturais (tipo de estrutura) e padrões prosódicos (acento e número de sílabas) estão envolvidos na formação de um *blend* em português. Entender como se dá a combinação desses fatores foi fundamental para entendermos como essas palavras são processadas, em uma etapa que envolveu alguns testes psicolinguísticos, que começaram a ser realizados tomando falantes do português brasileiro (PB) e do português europeu (PE). Nesta pesquisa, realizada na Universidade de Lisboa, em parceria com a Profa. Dra. Alina Villalva, realizamos um teste off-line de Associação de Palavras (TAP), em que perguntávamos aos falantes das duas línguas a primeira palavra que lhes vinha à mente quando liam um *blend*.

Os resultados desse primeiro TAP foram bastante reveladores. Nossa hipótese era a de que, quando havia uma estrutura do tipo CP (e.g., *impastor* <impostor e

pastor>) ou PC (e.g., *tristemunho* <triste + testemunho>), em que há palavras visíveis, os participantes responderiam com a palavra visível, contudo, isso não aconteceu. Houve uma preferência por parte dos participantes, em ambas as línguas, por respostas que envolviam a palavra clipada, isto é, a palavra-base que não estava totalmente visível no *blend*. Em estruturas do tipo PC foram cerca de 63% de respostas para a Base 2, e em estruturas do tipo CP foram 70% de respostas para a Base 1. Novos testes apontaram para o fato de que havia a atuação de um componente prosódico: a palavra clipada podia ser recuperada, principalmente quando o número de sílabas do *blend*, ou o acento, era o mesmo do *blend*, como em *delegata*, ou seja, uma das bases era parcialmente apagada, mas seu padrão silábico ou acentual era mantido no *blend*, possibilitando sua recuperação pelos participantes.

Também um teste online de Decisão Lexical, em que se contabiliza o tempo que o participante leva para ler uma palavra e responder se a palavra pode ser uma palavra do português ou não, foi realizado com falantes do PB e do PE. O tempo contabilizado desde o momento em que o participante lê a palavra até o momento em que dá a sua resposta é chamado de *tempo de reação*, e os resultados desse teste podem ser observados no Gráfico 1:



Gráfico 1 — Médias do tempo de reação no teste de Decisão Lexical para o PE e PB.

O Gráfico 1 mostra os resultados dos tempos de reação (TR) em milissegundos para cada uma das estruturas PC, CP e CC. Primeiramente, notamos que os *blends* com a estrutura CC tiveram um TR menor nas duas línguas, o que, para nós, naquele momento, foi curioso, pois a estrutura CC é a estrutura em que nenhuma das bases do *blend* é visível. Nossa hipótese era a de que não tendo bases visíveis, o TR seria maior. Também é possível perceber, pelas linhas do gráfico, que há uma certa homogeneidade no processamento dos *blends*, por meio dos TR, nas duas línguas. Embora os participantes brasileiros tenham médias de TR mais rápidas, as categorias tiveram um comportamento bastante semelhante.

Recentemente, com a ascensão das questões ligadas à Inteligência Artificial, realizei uma pesquisa, juntamente com Thuany Teixeira de Figueiredo (Unicamp) e Rafael Beraldo (Impa Tech) sobre o comportamento dos Modelos de Linguagem de Grande Escala (LLMs) em relação à interpretação dos *blends*. Alguns resultados preliminares mostraram que, embora os LLMs tenham menor taxa de acerto na interpretação dos *blends* do que os humanos, há uma semelhança com os humanos no padrão de respostas, isto é, para uma estrutura como PC, os LLMs deram como resposta, na maioria das vezes, a base clipada. Também notamos um avanço na taxa de acertos de 15% para o GPT-3.5 e 47,5% para o GPT-4. Em breve, traremos novas notícias sobre os resultados dessa pesquisa.

Por fim, há um problema importante sobre o estudo dos *blends* que esbarra na questão de se *blends* podem ser considerados um tipo de nome composto. Essa é uma questão aberta, uma vez que há autores, como Sandmann (1991) e Marangoni Junior (2021), que aproximam os *blends* dos compostos, ao considerarem que *blends* podem ser um subtipo de composição, e outros autores, como Gonçalves (2003), Andrade e Rondinini (2016) e Braga e Pacheco (2024), que defendem que os *blends* não são compostos. Nesse ponto, vale a pena mencionar que o *blending* e os *blends* são ignorados pelos livros didáticos e documentos como a Base Nacional Comum Curricular (BNCC). O próprio processo de composição é relegado a classificações que exploram apenas aspectos fonológicos como, por exemplo, a justaposição e a aglutinação, enquanto aspectos da estrutura e da modificação nessas formações são deixados de lado, sem contar a utilização de exemplos que não fazem parte do cotidiano dos alunos como, por exemplo, “fidalgo” e “pernalta”. Como já mencionei anteriormente, os *blends* são frequentes no cotidiano dos falantes do português e, por todas as propriedades elencadas, deveriam ser explorados nas aulas da educação básica no ensino de morfologia e para a conscientização morfológica dos alunos.

Saiba mais

MINUSSI, Rafael Dias; NÓBREGA, Vitor Augusto. A interface sintaxe-pragmática na formação de palavras: avaliando os pontos de acesso da Enciclopédia na arquitetura da gramática. **Veredas**, Juiz de Fora, v. 18, n. 1, pp. 161–184, 2014.

MINUSSI, Rafael Dias; VILLALVA, Alina Maria Santos Mártires. Reconhecimento e acesso lexical dos blends em português europeu e português brasileiro. **Todas as Letras – Revista de Língua e Literatura**, São Paulo, v. 22, n. 1, pp. 1–14, 2020.

Referências bibliográficas

ANDRADE, Katia Emmerick; RONDININI, Roberto Botelho. Cruzamento vocabular: um subtipo da composição? **DELTA: Documentação e Estudos em Linguística Teórica e Aplicada**, São Paulo, v. 32, n. 4, pp. 861–887, 2016.

BRAGA, Emerson Viana; PACHECO, Vera. Algumas considerações sobre o status dos blends no léxico do português brasileiro. **Revista do GEL**, v. 21, n. 1, pp. 54–78, 2024.

GONÇALVES, Carlos Alexandre. *Blends* lexicais em português: não-concatenatividade e correspondência. **Veredas**, Juiz de Fora, v. 7, n. 1, pp. 149–167, jan./dez. 2003.

MARANGONI JUNIOR, César Elidio. **A blenditividade na formação de palavras**: a derivação dos *blends* na interface entre morfologia, fonologia e pragmática. 2021. Dissertação (Mestrado em Semiótica e Linguística Geral) — Faculdade de Filosofia, Letras e Ciência Humanas, Universidade de São Paulo, São Paulo, 2021.

SANDMANN, Antônio José. **Morfologia Geral**. São Paulo: Contexto, 1991.

VILLALVA, Alina; MINUSSI, Rafael Dias. Description and analysis of a Portuguese blend corpus. **Corpus**, n. 23, 2022.