

O uso da análise de clusters na identificação de padrões de transitividade linguística

The use of cluster analysis for identification of linguistic transitivity patterns

Marcus Lepesqueur
Universidade Federal de Minas Gerais
marcus.le@gmail.com

Ilka Afonso Reis
Universidade Federal de Minas Gerais
ilka@est.ufmg.br

Resumo

Este trabalho visa a apresentar a técnica de agrupamento hierárquica para análise de padrões semânticos e sintáticos da transitividade no nível oracional. Partindo de uma perspectiva empírica e baseando-se em dados reais da língua em uso, esse tipo de metodologia se mostrou útil na investigação dos padrões linguísticos a que os falantes são expostos, chegando a resultados semelhantes a categorias teoricamente conhecidas. Em um processo de amostragem simples sem reposição, foram selecionadas 690 unidades oracionais de um *corpus* de 23 entrevistas orais. Essas unidades oracionais foram analisadas em termos de nove parâmetros de transitividade e de sua respectiva sintaxe oracional. O objetivo foi identificar grupos de orações que compartilham semelhanças em termos de um conjunto de traços semânticos e morfossintáticos. Os grupos encontrados revelam um tipo de significado protoconceptual das orações, que inclui traços aspectuais e actanciais que se correlacionam. Os resultados evidenciam três cenários micro-narrativos básicos sobre os quais se desenrola o evento expresso na oração.

Palavras chave

transitividade, análise de *cluster*, semântica, sintaxe oracional

Abstract

This paper aims to present a hierarchical clustering technique for the analysis of semantic and syntactic patterns of transitivity at clausal level. From an empirical and usage-based approach, this type of methodology has proved useful for the investigation of linguistic patterns to which speakers are exposed, reaching similar results found in theoretically categories. In a simple sampling procedure without replacement, 690 oral units were selected from a corpus of 23 oral interviews. These sentence units were analyzed in terms of nine transitivity parameters and their clausal syntax. The goal was to identify groups of sentences that share similarities in terms of this set

of traits. The groups found reveal a kind of protoconceptual meaning of the sentences, which includes correlated aspectual and actantial traits. The results show three basic micro-narrative scenarios on which the event expressed in clausal unfolds.

Keywords

transitivity, cluster analysis, semantics, clausal syntax

1 Introdução

A transitividade tem um papel central em grande parte das teorias linguísticas, principalmente pelo fato de que um número muito significativo de línguas apresenta uma estrutura formal —a morfossintaxe transitiva— cuja principal função é expressar um conjunto específico de propriedades semânticas (Næss, 2007). A questão principal em torno do fenômeno da transitividade é que as características semântica e sintática da transitividade tendem a covariar, sendo um fenômeno universal, ou ao menos *quasi* universal das línguas humanas. Givón (2001) aponta que, apesar das características transitivas de uma oração parecerem independentes, é um fato, na maioria das línguas, que as estruturas sintática e semântica da transitividade se sobrepõem, de forma que grande parte das orações semanticamente transitivas são também sintaticamente transitivas. De forma semelhante, Næss (2007) tenta demonstrar que, em muitas línguas, uma oração que é formalmente distinta da oração transitiva também se desvia dessa oração transitiva em termos das suas propriedades semânticas, ou seja, a escolha por uma estrutura linguística diferente reflete o desejo do falante de exprimir uma semântica diferente do protótipo da transitividade.

Mas apesar dos esforços das últimas décadas, a linguística contemporânea não encontrou uma explicação completa capaz de abraçar a complexidade desse fenômeno. Primeiro, porque os verbos variam as características da sua regência



em diferentes contextos de uso; segundo, porque as definições tradicionais da transitividade tratam, comumente, da mesma forma, elementos sintáticos e semânticos, que não apenas são distintos, mas que também interagem de forma complexa (Lucena & Cunha, 2012). Influenciada por fatores diacrônicos e sincrônicos, a morfossintaxe transitiva pode acomodar uma gama de valores semânticos não transitivos e, inversamente, a semântica transitiva pode ser expressa por mais de um padrão formal.

Por exemplo, no português do Brasil, como em muitas outras línguas, algumas orações sintaticamente transitivas não apresentam a semântica transitiva. Enquanto orações transitivas prototípicas como em (1) expressam ações, alguns verbos psicológicos transitivos, como em (2), têm significado estático, sendo considerados verbos de estado mental.

- (1) eu raspei a barba¹
 (2) ela sabe tudo

Na oração transitiva, normalmente temos, como sujeito sintático, o causador do evento e, como objeto, o participante afetado. No entanto, encontramos casos em que essas funções não são claras. Naess (2007) usa o termo “inversão do vetor semântico” para descrever uma relação semântica aparentemente invertida. Em (3), por exemplo, o sujeito sintático “Eu” não é exatamente o causador do evento, e o objeto sintático “ela” não é afetado pelo evento.

- (3) Eu vi ela

Tradicionalmente, a literatura linguística tenta associar a sintaxe transitiva à semântica de um sistema causal físico; no entanto, um número significativo de estruturas transitivas exibe relações não-físicas ou sem uma conexão direta do tipo causa-efeito. Por exemplo, “convidar” em (4a) e “chamar” em (4b), mesmo em estruturas transitivas, não expressam, necessariamente, um efeito direto no participante que se encontra na posição do objeto direto.

- (4) a. Dilma convidou eu
 b. Chamei ela

Com o objetivo de abordar essa complexa interface da sintaxe e semântica transitiva, este artigo investigou, em uma perspectiva empírica,

¹Os exemplos apresentados aqui foram retirados do *corpus*. Na transcrição foram respeitados alguns padrões de pronúncia, tais como a ausência da fricativa glotal surda (/r/) nos infinitivos verbais, ausência de morfema de plural e reduções como “tá”, para “está” ou “cê” para “você”.

alguns aspectos gramaticais das unidades oracionais no português do Brasil. Mais precisamente, esse trabalho visa a apresentar uma técnica hierárquica de agrupamento para analisar grupos de orações que compartilham aspectos semânticos e sintáticos da transitividade. Partindo de uma perspectiva empírica e baseando-se em dados reais da língua em uso, esse tipo de metodologia se mostrou útil na investigação dos padrões linguísticos a que os falantes são expostos, chegando a resultados semelhantes a categorias teoricamente conhecidas.

2 Parâmetros de Transitividade

Em um artigo seminal sobre o tema, Hopper & Thompson (1980) isolaram alguns dos componentes da noção de transitividade e estudaram a forma como eles são tipicamente codificados na gramática. A partir de um conjunto de evidências translinguísticas os autores propuseram um conjunto de parâmetros semânticos e morfossintáticos relacionados ao fenômeno da transitividade.

Hopper & Thompson (1980) argumentam que as gramáticas das línguas agrupam estes parâmetros em função de uma escala de transitividade, i.e., em uma mesma sentença, um traço morfossintático ou semântico obrigatório que marca alta transitividade tende a não ocorrer com outro que marca baixa transitividade.

Cada parâmetro proposto por estes autores², essencialmente, captura algum tipo de diferença entre as unidades oracionais. O primeiro parâmetro proposto, denominado Número de Participantes, distingue a particularidade de orações que aparecem sem objeto sintático, como (5a), daquelas que têm um ou mais objetos sintáticos, como (5b).

- (5) a. a intimação estourô
 b. eu disse pra ela do medo

O segundo, o terceiro e o quarto parâmetros da transitividade analisados nesta pesquisa são a Cinese, a Telicidade e a Pontualidade da predicação. Esses parâmetros fazem parte da noção mais geral de aspecto e referem-se à forma de se conceptualizar a estrutura temporal interna de uma determinada situação. Cinese refere-se à

²Nesta pesquisa, foram incluídos todos os parâmetros de transitividade com exceção da Individuação do objeto, que se refere a um conjunto variado de traços, o que inclui aspectos da referencialidade e definição/indefinição do objeto sintático. Este parâmetro depende de um tratamento teórico inconcluso e Hopper & Thompson (1980) o operacionalizaram em uma escala própria, distinta dos demais.

distinção entre predicacões que expressam ações, como em (6a) daquelas que expressam estados, como em (6b). Telicidade refere-se à presença ou à ausência do traço télico do evento, i.e., se o evento é conceptualizado como tendo um ponto de conclusão definido, como em (7a), ou sem este ponto de conclusão, como em (7b). Pontualidade distingue eventos pontuais, que não possuem fases intermediárias entre o seu início e o seu final, como em (8a), de eventos durativos, como em (8b).

- (6) a. minha mãe me ligô esses dia
b. os médico daqui eles são muito bom
- (7) a. tinha que fazê o relatório
b. aí eu andava bastante
- (8) a. ele morreu
b. reformô minha casa toda

Com o parâmetro Modalidade, Hopper & Thompson (1980) fazem a distinção entre o modo *realis* e o *irrealis* do evento, marcando a oposição entre a forma indicativa e formas não assertivas, tais como o subjuntivo, o condicional, o hipotético etc. O parâmetro da Modalidade distingue, portanto, o grau de realização do evento linguístico expresso. Também o parâmetro Polaridade, definido como a distinção entre a forma negativa e a forma afirmativa da oração, relaciona-se com o grau de realização do evento.

Os parâmetros Agentividade e Intencionalidade referem-se ao grau de envolvimento do argumento externo na atividade expressa pelo verbo. A Agentividade refere-se ao elemento desencadeador do processo e Intencionalidade refere-se ao sentido de volição. Em (9) encontra-se um exemplo de unidade oracional com argumento externo não agentivos e não volitivos e em 10 o argumento externo agentivo e volitivo.

- (9) a mente dele é de criança
- (10) e eu falei pra ela

Por fim, a noção semântica de Afetação, último parâmetro analisado aqui, é tradicionalmente um dos critérios essenciais na definição de transitividade. Muitas línguas mostram um padrão de se codificarem argumentos fortemente afetados pelo evento verbal como objetos de construções transitivas, e argumentos não afetados, ou menos afetados, em outras posições sintáticas (Lepesqueur, 2017). Em (12) encontra-se um exemplo típico da presença do parâmetro Afetação na construção transitiva.

- (11) eu podia amputá sua perna

No modelo de Hopper & Thompson (1980) a transitividade passou a ser definida não como uma característica do elemento verbal, mas como um conjunto de componentes ligados à unidade oracional que se relacionam de maneiras específicas. Essa mudança de perspectiva alimentou uma série de pesquisas que investigaram tanto a maneira como as línguas codificam formalmente os parâmetros da transitividade, quanto as motivações semânticas e pragmáticas da variação na morfossintaxe transitiva.

No português do Brasil, Lepesqueur (2017) mostrou que apenas um dos parâmetros propostos por Hopper & Thompson (1980), a saber, a Afetação do objeto sintático, é um preditor positivo, estatisticamente significativo, da sintaxe transitiva. Dito de outra forma, a presença da afetação do objeto na oração é um indicador de alta probabilidade da ocorrência da estrutura oracional transitiva. Os demais parâmetros de transitividade encontram-se distribuídos de maneira mais ou menos homogênea entre todas as estruturas oracionais, não compondo elementos distintivos da sintaxe transitiva. O autor sugere ainda que certos parâmetros, especialmente a Telicidade, podem estar associadas a padrões sintáticos não-transitivos. Isso aponta para certas particularidades da organização da estrutura transitiva no português e sugerem possibilidades de se repensar a associação entre a semântica e a sintaxe transitiva.

Como os parâmetros de transitividade, independentemente da estrutura da unidade oracional, se agrupam no português do Brasil? Existe apenas uma semântica não transitiva, ou podemos esperar diferentes padrões semânticos fora da transitividade? Partindo desse conjunto de questões, esse trabalho visa a apresentar uma metodologia estatística capaz de identificar padrões semânticos e sintáticos da oração, chegando a resultados semelhantes a categorias teoricamente conhecidas. O objetivo foi identificar grupos de orações que compartilham semelhanças em termos de um conjunto de traços morfossintáticos e semânticos ligados à transitividade. Esperamos que os resultados apresentados aqui possam elucidar as regularidades sintático-semânticas às quais os falantes estão expostos e sobre as quais emergem os fenômenos gramaticais.

3 Metodologia

Uma das principais dificuldades para a compreensão da transitividade é que estamos lidando em um campo de interface entre a estrutura formal e

a estrutura conceptual³. Apesar dos avanços recentes da linguística sobre a natureza dessa articulação, restam ainda muitas questões a respeito da maneira através da qual um item lexical se integra em uma sintaxe —e, mais ainda, em uma estrutura macrotextual e discursiva— e como isso pode produzir efeitos de significado.

Por consequência, a compreensão da transitividade depende, antes de tudo, de um tratamento de dois eixos distintos entre si: um eixo essencialmente semântico e outro essencialmente sintático. Por fim, além da descrição desses dois eixos, é preciso um modelo linguístico, talvez mais especificamente semiótico, que explique a maneira complexa e particular através da qual a sintaxe e a semântica transitiva interagem.

Em uma análise de interface entre sintaxe e semântica, o caminho tradicional de investigação da transitividade tem sido agrupar padrões morfossintáticos a fim de se analisar uma estrutura semântica subjacente. Assim, por exemplo, pode-se distinguir a estrutura formal transitiva da intransitiva para, em seguida, tentar-se identificarem as diferenças semânticas nesses grupos. Mas o caminho inverso também é possível: primeiro identificar grupos de orações semanticamente semelhantes e posteriormente analisar a relação desse grupo com padrões formais da língua.

Por diversas razões, a primeira opção tem sido o caminho canônico de investigação. Uma das principais questões é o fato da língua agrupar uma quantidade a princípio ilimitada de informações conceptuais em um número relativamente limitado de estruturas e regras gramaticais. Isso torna mais fácil agrupar as unidades oracionais a partir das suas características formais, que são em um número relativamente reduzido, do que agrupá-las a partir das suas distinções conceptuais. Desta perspectiva, vários teóricos têm tentado analisar padrões gramaticais (morfossintáticos e lexicais) buscando inferir uma estrutura conceptual subjacente. Este é o raciocínio básico por trás dos trabalhos de Hopper & Thompson (1980), Givón (2001) ou Næss (2007).

³Aqui utilizamos o termo *conceptual*, escrito com p, para destacar o caráter processual da estrutura semântica. Em geral, os teóricos da Linguística Cognitiva têm utilizado o termo *conceptualization* (traduzido normalmente como *conceptualização*) para se referir ao processo de construção de significado, destacando sua natureza dinâmica e processual. A *conceptualização* tem sido descrita como um processo imagético (em oposição à noção tradicional de estruturas proposicionais), interativo (porque envolve processos de negociação e interação entre os interlocutores), e imaginativo (porque envolve processos de simulação e mesclagens conceituais) (Broccias, 2013).

Um caminho alternativo é tomar a estrutura conceptual como um dado, perceptível pelos falantes, a fim de, posteriormente, estabelecerem-se relações simbólicas com a estrutura formal da língua. A proposta de Halliday et al. (2014), que compreende o sistema da transitividade como uma função gramatical organizadora, com seus próprios modelos e esquemas, é um exemplo da tentativa de focalizar, inicialmente, a maneira como a informação conceptual é estruturada para, posteriormente, identificar sua manifestação formal.

Este trabalho parte desta última via e tem como objetivo identificar grupos de orações que compartilham semelhanças em termos do conjunto de parâmetros de transitividade como um todo, de maneira parcialmente⁴ independente da estrutura formal da oração e, posteriormente, tentar estabelecer uma relação entre os parâmetros e a estrutura sintática oracional do português do Brasil. Buscamos identificar agrupamentos naturais de unidades oracionais (grupos de orações que compartilham semelhanças em termos dos seus parâmetros) a partir de um conjunto de técnicas estatísticas de agrupamento. Essa metodologia mostrou-se capaz de analisar, empiricamente, traços semânticos ou morfossintáticos em dados reais da língua em uso, chegando a resultados semelhantes àqueles esperados teoricamente

3.1 Composição do *corpus*

O *corpus* desta pesquisa é composto de relatos orais produzidos por 23 participantes, publicados em Lapesqueur (2017)⁵. As narrativas orais produzidas por esses participantes foram gravadas e transcritas. Para facilitar a importação e o tratamento dos dados pelo programa computacional de análise estatística, cada linha da transcrição contém o trecho correspondente

⁴Dizemos parcialmente porque os parâmetros não são puramente semânticos. Por exemplo, o parâmetro Afetação refere-se a uma distinção semântica que ocorre em uma certa posição sintática, a saber, a posição de objeto. Mas esse objeto pode ser, a princípio, preposicionado ou não, ou fazer parte de uma estrutura sintática transitiva ou bitransitiva.

⁵A pesquisa de Lapesqueur (2017) teve o objetivo de investigar o fenômeno da transitividade em uma população clínica. Parte do *corpus*, portanto, é composto de entrevistas produzidas por pacientes com diagnóstico de esquizofrenia paranoide. O referido trabalho não identificou algum tipo de correlação especial intra-parâmetros na população clínica, apenas a maior probabilidade de ocorrer o parâmetro Afetação na fala dos pacientes. Uma vez que trata-se de um parâmetro pouco frequente no *corpus*, não há evidências de que os agrupamentos apresentados neste trabalho não possam ser generalizados.

a uma única unidade oracional, definida como uma predicação centralizada pela unidade verbal. As transcrições foram realizadas usando-se as convenções ortográficas, sem, no entanto, dar atenção especial às questões fonéticas, uma vez que não possuem relevância para a pesquisa.

Do total de 7939 unidades oracionais da transcrição, 5690 fizeram parte da análise, uma vez excluídos trechos do entrevistador, unidades oracionais abandonadas ou parcialmente incompreensíveis, expressões idiomáticas e estruturas não sentenciais.

Em um processo de amostragem simples sem reposição, foram selecionadas 690 unidades oracionais (30 por participante), analisadas em termos dos parâmetros de transitividade e sua sintaxe oracional. Os *corpus* é original da pesquisa de Lepesqueur (2017), que definiu o tamanho da amostra respeitando o número mínimo de observações sugeridas por Hair et al. (2009) para análise de regressão logística. O autor também considerou um desenho experimental balanceado em termos do número de observações por participantes. Os dados foram organizados em uma planilha eletrônica de forma a conter, para cada unidade oracional observada, a classificação dos parâmetros de Hopper & Thompson (1980) e Thompson & Hopper (2001), em termos de alta (1) ou baixa (0) transitividade⁶. Posteriormente, esses dados foram analisados no ambiente de programação estatística R (R Development Core Team, 2017).

3.2 Análise de clusters

A análise de *cluster* (também conhecida como análise de conglomerado ou de agrupamentos) é um conjunto de algoritmos e de técnicas analíticas multivariadas que visa a agrupar os elementos de uma amostra ou população a partir da similaridade desses elementos quando os comparamos em uma série de variáveis (Mingoti, 2017). O objetivo desse tipo de técnica é realizar agrupamentos que maximizem as semelhanças entre observações que pertençam a um mesmo

grupo, o que torna o grupo mais homogêneo, ao mesmo tempo em que minimizem as semelhanças entre grupos diferentes, o que torna os grupos heterogêneos entre si. No campo dos estudos linguísticos a análise de agrupamentos tem sido utilizada para descrever uma ampla gama de fenômenos que vão desde diferenças dialetais até polissemias (Divjak & Fieller, 2014).

Uma questão central desse tipo de análise refere-se à métrica utilizada para se decidir o grau de similaridade (ou inversamente, de dissimilaridade) entre os elementos observados. No caso de variáveis qualitativas, tais como os parâmetros binários de transitividade analisados aqui, foi utilizado o coeficiente de concordância simples (s_{ij}) (Sokal & Sneath, 1963). Trata-se de um coeficiente simétrico, ou seja, que considera o mesmo peso para as concordâncias positivas ou negativas. O coeficiente é calculado pela soma do número total de concordâncias entre os atributos dos elementos i e j , dividido pelo número total de atributos.

$$s_{ij} = \frac{\text{Número de atributos concordantes}}{\text{Número total de atributos}}$$

O valor de s_{ij} pode variar entre 0 e 1. O coeficiente s_{ij} foi calculado, para cada $i \neq j$, através do coeficiente geral de Gower (1971) que permite integrar também, se necessário, variáveis quantitativas ou ordinais.

Considerando o coeficiente de similaridade s_{ij} , a matriz de dissimilaridade dos dados será composta pelo índice de dissimilaridade d_{ij} , calculado pelo complementar de s_{ij} para cada par de orações do *corpus*.

A partir dessa matriz de dissimilaridade foram utilizadas técnicas hierárquicas de agrupamento para encontrar a melhor partição dos dados. Optamos pelo uso das técnicas hierárquicas, em uma análise exploratória, uma vez que não temos um número pré-estabelecido de grupos e buscamos identificar uma estrutura natural dos dados. As análises foram conduzidas utilizando-se tanto técnicas hierárquicas aglomerativas⁷ quanto a divisivas⁸ (Rousseeuw & Kaufman, 1990).

A técnica aglomerativa começa com n grupos, sendo n o número de elementos no banco de dados. Cada observação é separada em um *cluster* específico e o algoritmo de agrupamento tenta encontrar os valores mais semelhantes para formar os grupos. Inversamente, a técnica divisiva assume inicialmente todos os elementos em um único grupo e inicia a divisão dos elementos

⁶Essa análise foi feita manualmente a partir dos critérios descritos em Lepesqueur (2017). O coeficiente de Kappa sugere uma concordância substancial entre dois avaliadores especialistas, previamente treinados ($k=0.74$, $p < 0.05$). Não existe hoje uma boa abordagem para automatizar a avaliação dos parâmetros de transitividade. Isso depende, antes de tudo, de uma maneira eficiente de tratar computacionalmente valores semânticos ligados especialmente ao elemento verbal. A esse respeito, um dos projetos pioneiros é o FrameNet, idealizado por Chales Fillmore, no campo da Semântica de *Frames*. Para mais detalhes ver sobre o projeto ver <https://framenet.icsi.berkeley.edu>

⁷Função *hclust*, do pacote *stats* do software R.

⁸Função *diana*, do pacote *cluster* do software R.

mais distantes em grupos diferentes. A similaridade entre dois conglomerados foi definida pelo método de ligação completa, ou seja, a partir da comparação da maior distância entre os pontos de dois grupos. Esse método tende a formar grupos mais compactos e sem a tendência de longas cadeias⁹.

Para decidir sobre o número K de grupos da partição final dos dados analisados, utilizamos algumas medidas de avaliação da qualidade dos agrupamentos, analisando tanto a compactidade (a máxima similaridade intra-grupo) quanto a separabilidade (a mínima similaridade entre grupos).

Inicialmente utilizamos duas medidas de avaliação de todas as partições de 2 a 30 *clusters*¹⁰, tanto na técnica aglomerativa quanto na divisiva. A primeira medida foi uma generalização da soma de quadrados dos desvios intra-*cluster* (tipicamente utilizada na métrica euclidiana) e a segunda medida a largura de silhueta (Rousseeuw, 1987).

A soma de quadrados dos desvios intra-*cluster* (SQ_k) é uma estimativa da compactidade de um dado *cluster* k e se refere, aqui, à metade da soma dos quadrados das dissimilaridades intra-*cluster* dividido pelo tamanho do *cluster*. SQ_k é definido como:

$$SQ_k = \frac{1}{2n_k} \sum_{i,j \in C_k} d_{ij}^2$$

onde n_k é o número de elementos no *cluster* C_k e d_{ij} é o valor da dissimilaridade entre o elemento i e j do *cluster* C_k . Quanto maior o valor de SQ_k , menor será a compactidade do *cluster* k . A soma de quadrados dos desvios intra-*cluster* é uma medida particular do *cluster* k . Na partição final com K *clusters*, cada um desses *clusters* apresenta um valor próprio de SQ_k . Para uma dada partição final, a soma de quadrados dos desvios intra-*cluster* desta partição é dada pela média dos valores de SQ_k de todos os K *clusters*.

A largura média de silhueta (L) oferece uma estimativa da separabilidade dos agrupamentos ao comparar a similaridade de uma observação amostral com as demais observações do próprio *cluster* e do seu vizinho mais próximo. A largura média de silhueta é calculada a partir do coeficiente de silhueta (S_i) da observação amostral i :

⁹Que ocorre quando um *cluster* incorpora, a cada iteração, um único elemento próximo.

¹⁰A princípio, não esperamos que haja mais de 30 grupos teoricamente importantes para explicar o fenômeno da transitividade. Mas não se trata de uma restrição da técnica. Ainda que computacionalmente demorado, é possível analisar até $n-1$ agrupamentos, sendo n é o número de observações no banco de dados.

$$S_i = \frac{b_i - a_i}{\text{MAX}(a_i, b_i)}$$

onde a_i é a média da dissimilaridade (d_{ij}) da observação amostral i com todos os membros do *cluster* ao qual pertence e b_i , a dissimilaridade (d_{ij}) mínima da observação i com todos os demais dados que não pertencem ao seu *cluster*.

O coeficiente de silhueta S_i varia no intervalo de $[-1, 1]$ e se aproxima de -1 quando o elemento i está, em média, mais próximo dos elementos de um *cluster* vizinho do que dos elementos do seu próprio *cluster* (caso em que $b_i < a_i$). O coeficiente aproxima-se de 0 na medida em que b_i seja semelhante a a_i , sugerindo que o elemento i encontra-se em um ponto intermediário entre dois *clusters*. O coeficiente aproxima-se de 1 quando o elemento i está em média mais próximo dos elementos do próprio *cluster* do que do *cluster* vizinho (caso em que $b_i > a_i$).

A largura de silhueta (S_k) de um *cluster* k é dada pela média dos coeficientes de silhueta de todas as observações pertencentes ao *cluster* k . Por suas vez a largura média de silhueta (L) da partição foi calculado pela média dos valores de S_k dos *clusters* S_k que compõem aquela partição.

3.3 Escolha da melhor técnica de agrupamento

A partir dos índices apresentados, definimos a melhor partição obtida na técnica aglomerativa e a melhor partição obtida na técnica divisiva. Para auxiliar na comparação entre essas duas partições finais, iniciamos um segundo passo de análise das medidas da qualidade da partição a partir do Índice Dunn2 (Halkidi et al., 2001) e Índice WB. Ambos os índices são calculados a partir da dissimilaridades intra-*cluster* $d(C_k)$ e a dissimilaridade entre *clusters* $d(C_k, C_l)$.

Quanto menor as dissimilaridades intra-*cluster*, maior a compactidade da partição. A dissimilaridade intra-*cluster* do *cluster* k é dado por:

$$d(C_k) = \frac{2}{n_k(n_k - 1)} \sum_{i \in C_k, j \in C_k} d_{ij}$$

Quanto maior as dissimilaridades entre *clusters*, maior a separabilidade da partição final. A dissimilaridade entre o *cluster* k e l é dado por:

$$d(C_k, C_l) = \frac{1}{n_k n_l} \sum_{i \in C_k, j \in C_l} d_{ij}$$

O Índice WB (*whitin/between*), I_{wb} , é calculado pela razão entre as médias de $d(C_k)$ e

$d(C_k, C_l)$ para todos os *clusters* da partição final. Quanto menor o índice WB, melhor a relação entre compacidade (numerador) e separabilidade (denominador).

O Índice Dunn2 é dado pela razão entre a menor dissimilaridade entre dois *clusters* e a maior dissimilaridade intra-*cluster* da partição final. Quanto maior o índice, melhor a relação entre a separabilidade (numerador) e a compacidade (denominador).

A partição final, depois de comparadas as técnicas aglomerativa e divisiva, foi representada graficamente utilizando a técnica de escalonamento multidimensional (MDS). O método MDS faz a decomposição espectral de uma matriz relacionada à matriz de dissimilaridade entre os elementos amostrais. Assim, ao se construir novas dimensões e grafar seus valores num gráfico de dispersão, conserva-se aproximadamente as dissimilaridades que os elementos amostrais apresentam entre si. Em suma, essa técnica permite representar espacialmente a matriz de dissimilaridade dos elementos sintetizando essa matriz em um certo número de componentes utilizadas como coordenadas de um gráfico de percepção. Neste gráfico, as relações geométricas correspondem, de maneira aproximada, às relações de dissimilaridade dos dados observados (Mingoti, 2017).

4 Resultados e discussão

As unidades oracionais do *corpus* foram analisadas em termos dos 9 parâmetros propostos por Hopper & Thompson (1980), sendo cada um destes parâmetros caracterizado como de alta (1) ou baixa (0) transitividade. A Figura 1¹¹ mostra a distribuição, no *corpus*, da frequência absoluta dos parâmetros, segundo o seu grau de transitividade. Uma vez que cada parâmetro soma 690 observações (tamanho da amostra), o gráfico representa também, visualmente, a proporção relativa dos traços de baixa e alta transitividade.

Alguns traços de transitividade são especialmente raros na amostra analisada: a afetação do objeto sintático (Afetação=Alta) ocorre em menos de 10% do *corpus* e a baixa polaridade da oração (Polaridade=Baixa) em apenas 12,6%. A maior proporção de orações afirmativas é, provavelmente, uma consequência do gênero textual do *corpus*, composto por entrevistas orais.

¹¹Na figura os parâmetros de transitividade foram abreviados e são apresentados na seguinte ordem: Telicidade (Telicid.), Pontualidade (Pont.), Polaridade (Pol.), Participante (Part.), Modalidade (Mod.), Intencionalidade (Intenc.), Cinese (Cin.), Agentividade (Agent.), Afetação (Afet.)

De maneira geral, a presença de traços de baixa transitividade são mais comuns na amostra do que traços de alta transitividade¹². Esta característica já era esperada uma vez que a bibliografia especializada tem afirmado que o gênero conversação tende a ser de baixa transitividade, como sugerem Thompson & Hopper (2001), para o inglês, Rozas (2004), para o espanhol, Shahrokhi & Lotfi (2012), para o persa, e Lima (2013), para o português. Bois (2003), analisando a preferência no discurso pelo uso de certas configurações sintáticas, mostrou que, em diversas línguas (a saber, Hebrew, Sakapultek, Papago, Inglês e Goonyandi), 50 a 62% das unidades oracionais não possuem nenhum argumento nominal. De maneira geral, as orações de baixa transitividade parecem ser mais úteis no contexto de comunicação interpessoal e de aspectos subjetivos do que as orações de alta transitividade (Rozas, 2004).

4.1 Análise de agrupamento por técnica hierárquica

Para a investigação do melhor agrupamento dos dados, iniciamos com a análise do nível de fusão dos aglomerados. À medida que o número de *clusters* na partição aumenta, a média da dissimilaridade intra-*cluster* decresce. As Figuras 2 e 3, mostram os valores da média de SQ_k de todos os *clusters* que formam cada partição, tanto na técnica hierárquica aglomerativa quanto na divisiva.

Buscamos identificar nos gráficos mudanças significativas (“quedas bruscas”) na média da soma de quadrados da dissimilaridade, o que representa ganhos importantes na homogeneidade ou compacidade dos agrupamentos. No uso da técnica aglomerativa, na Figura 2, destaca-se o ganho de consistência interna na partição $K = 13$. No uso da técnica divisiva, na Figura 3, a queda significativa na soma de quadrado dos desvios ocorre na partição $K = 3$.

As Figuras 4 e 5, a seguir, mostram a média das larguras de silhueta de todos os *clusters* que compõem as partições com 2 a 30 grupos. Buscamos aqui as partições com maiores médias da largura de silhueta, o que representa maior separabilidade entre grupos vizinhos. Com o uso da técnica aglomerativa, o salto na média da largura de silhueta ocorre novamente no agrupamento de $K = 13$, com ganhos poucos significativos depois

¹²Especialmente se retiramos o parâmetro Polaridade, que deixou de ser considerado relevante na descrição do fenômeno da transitividade em Thompson & Hopper (2001).

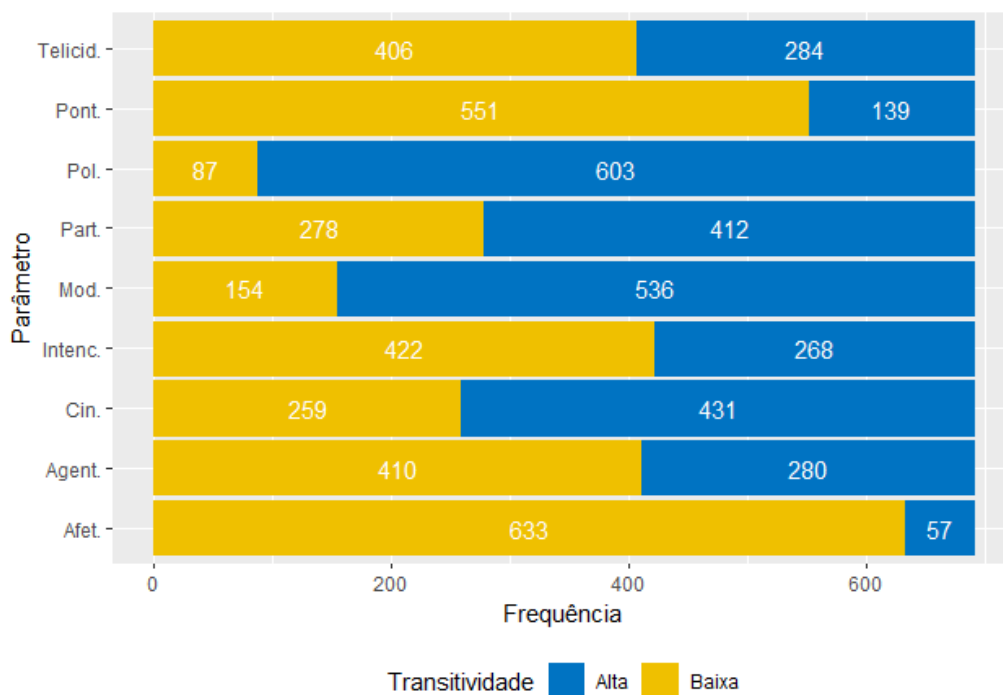


Figura 1: Gráfico de barras da frequência absoluta dos parâmetros de transitividade.

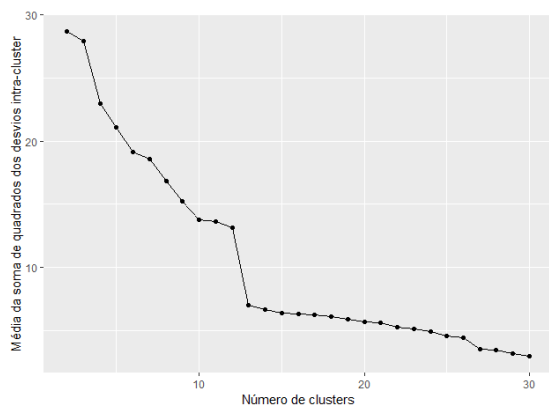


Figura 2: Média da soma de quadrados da dissimilaridade intra-cluster na Técnica Aglomerativa

dessa partição. Na técnica divisiva, o pico ocorre na partição com $K = 3$.

As duas medidas de avaliação da qualidade dos agrupamentos sugerem, portanto, uma partição com $K=3$, no uso da técnica divisiva, ou com $K=13$, no uso da técnica aglomerativa. A Tabela 1 apresenta a comparação das duas partições através de outros índices de avaliação da qualidade dos agrupamentos.

A Tabela 1 mostra um melhor desempenho da partição $K = 3$ (técnica divisiva) nos índices Dunn2 e média da largura de silhueta (sendo ambos os índices uma estimativa da relação entre compacidade e separabilidade), além do me-

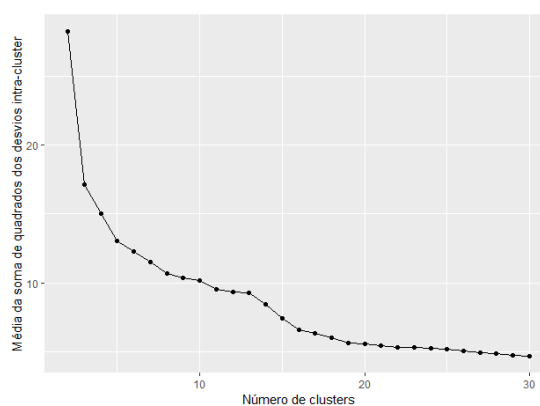


Figura 3: Média da soma de quadrados da dissimilaridade intra-cluster na Técnica Divisiva

lhor desempenho na média da dissimilaridade entre clusters (Média de $d(C_k, C_l)$). Apesar da partição final com $K = 13$ apresentar menor dissimilaridade intra-clusters (Média de $d(C_k)$), e consequentemente, melhor desempenho na razão (I_{ub}), esse ganho não acompanha a perda em parcimônia no uso de um número tão grande de grupos. Optou-se, portanto, pela partição final com $K = 3$, utilizando-se a técnica hierárquica divisiva.

Utilizando a técnica de Escalonamento Multidimensional (MDS), é possível representar espacialmente, em um gráfico de percepção, a matriz de dissimilaridade dos elementos e a partição final dos dados. A Figura 6 representa as observações

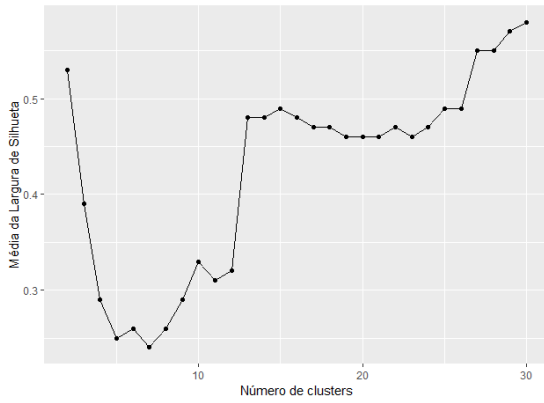


Figura 4: Média da Largura de silhueta na Técnica Aglomerativa

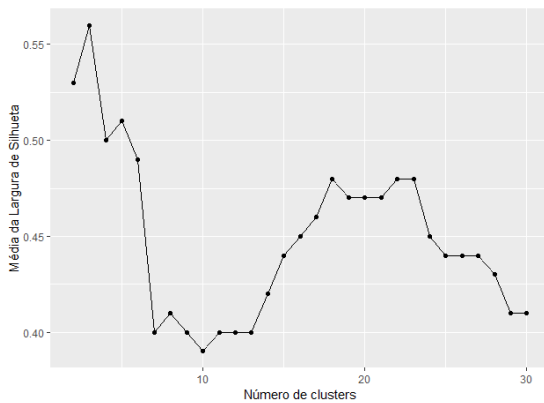


Figura 5: Média da Largura de silhueta na Técnica Divisiva

e o agrupamento final de maneira que a distância entre os pontos corresponde aproximadamente à dissimilaridade entre as observações.

A proporção da variância explicada pelas duas dimensões obtidas através do escalonamento é de 0,47. É importante notar que não pretendemos aqui realizar a análise de agrupamento a partir do MDS, mas apenas representar graficamente a distribuição dos *clusters*. O gráfico permite visualizar bem a correspondência entre a partição

Número de clusters	K=3	K=13
Média de SQ_k	17,16	7,01
Média de $d(C_k)$	0,18	0,09
Média de $d(C_k, C_l)$	0,50	0,43
I_{wb}	0,37	0,22
Índice Dunn2	1,75	0,93
Média da largura de silhueta	0,56	0,48

Tabela 1: Medidas de avaliação da qualidade dos agrupamentos (k=3, técnica divisiva; k=13, técnica aglomerativa).

final obtida no uso da técnica divisiva e a representação espacial das observações.

Para melhor identificar as características típicas de cada *cluster*, apresentamos também a frequência relativa das variáveis em cada um deles. A Figura 7 mostra a frequência relativa de ocorrências dos traços de transitividade em cada grupo da técnica divisiva $K = 3$, destacando em verde quando o traço ocorre em aproximadamente 100% das unidades oracionais do grupo em questão e, em amarelo, quando os parâmetros correm em aproximadamente 0% das unidades oracionais pertencentes àquele grupo. A ordem de apresentação das categorias nesse gráfico foi escolhida de modo a facilitar a visualização dos conjuntos de traços mais frequentes (blocos em verde) e os menos frequentes (blocos em amarelo) em cada grupo.

Percebe-se que os parâmetros Polaridade (negativa e afirmativa) e Modalidade (*realis* e *irrealis*) são relativamente distribuídos de forma homogênea entre os grupos. Os demais parâmetros se agrupam de forma bem definida, mostrando um padrão semântico específico de cada *cluster*, representado nos blocos em verde e amarelo.

O *cluster* 3 possui uma estrutura aspectual bem definida. Por estrutura aspectual, entendemos as diferenças da estrutura temporal interna, não relacionais, do evento expresso na oração (Comrie, 1976). Este grupo apresenta predicados que expressam estados (parâmetro Chinês=Baixa), sendo durativos (parâmetro Pontualidade=Baixa) e atéllicos (parâmetro Telicidade=Baixa). Eles são igualmente não agentivos e não intencionais (parâmetros agentividade e intencionalidade=Baixa). Em (12) encontramos um exemplo prototípico desse *cluster*.

Distintamente, os *clusters* 1 e 2 expressam ações (parâmetro Chinês=Alta). O *cluster* 1 expressa eventos não-agentivos e não-intencionais (parâmetros agentividade e intencionalidade=Baixa), tipicamente pontuais (parâmetro Pontualidade=Alta), enquanto o *cluster* 2 expressa eventos tipicamente agentivos, intencionais (parâmetros agentividade e intencionalidade=Alta) e durativos (parâmetro Pontualidade=Alta), podendo ou não apresentar um ponto télico (parâmetro Telicidade=Indefinido). Os exemplos (13) e (14) são prototípicos do *cluster* 1 e 2 respectivamente.

(12) ele é casado

(13) depois a intimação estourô

(14) eu tratava das criação

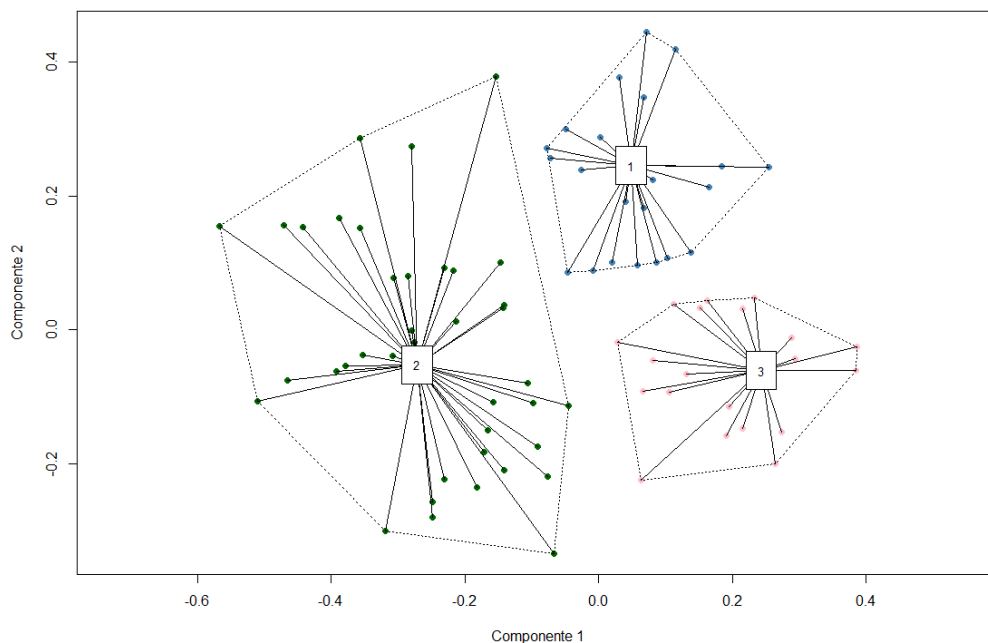


Figura 6: Gráfico de Percepção – Técnica divisiva $k = 3$.

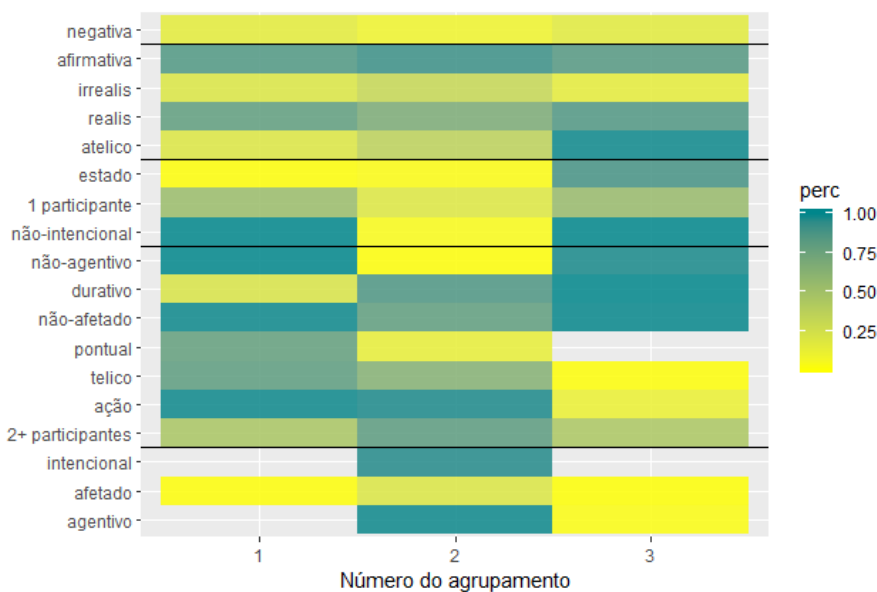


Figura 7: Frequência relativa dos parâmetros em cada *cluster* considerando a técnica divisiva.

4.2 Relação entre os agrupamentos e a estrutura sintática da oração

Especialmente no âmbito da Linguística Cognitiva, um dos conceitos chaves para a compreensão da estrutura semântica é a categorização. O processo de aquisição da linguagem envolve não apenas aprender quais categorias são relevantes para nós, em nosso ambiente, mas também aprender um número limitado de estruturas e regras gramaticais utilizadas para se expressar um

número ilimitado de experiências (Divjak & Feller, 2014).

A categorização é o resultado de uma capacidade cognitiva humana geral de realizar abstrações e reconhecer um núcleo comum de aspectos da experiência corpórea e social. Contrariamente à visão clássica que compreende os conceitos como representações de estados de um mundo objetivo e, portanto, não sujeitos à experiência subjetiva, estudos empíricos têm mostrado que

os conceitos são definidos e compreendidos dentro de um quadro conceitual que depende da natureza da experiência humana. Esta concepção denominada de actuação (*enaction*) ou corporeidade (*embodiment*) foi especialmente tratada por Johnson (2013) e por Varela et al. (1991), dentro das Ciências Cognitivas, e se resume na afirmação de que a cognição não pode ser compreendida fora de nossa história social e de ações corporalizadas. Por ação corporalizada, entende-se, primeiro, que a nossa cognição é inseparável da forma como experienciamos processos sensoriais e motores (percepção e ação) decorrentes de termos um corpo como o nosso e, segundo, que essa experiência encontra-se mergulhada em um contexto biológico, psicológico e cultural mais abrangente. A experiência envolve padrões recorrentes, ou *gestalts*, no sentido de uma organização coerente, que são fundamentais para o processo de significação e estão na origem de certos pontos de referência do nosso sistema conceitual (Johnson, 2013) (Lakoff, 2008). Em resumo, nosso sistema conceitual ancora-se em certos padrões de interação sensório-motoras, que servem de base para a significação.

O conjunto dos dados analisados neste trabalho revela que as características semânticas da transitividade podem ser agrupadas em padrões relativamente bem definidos em termos de suas características. Esses grupos parecem instanciar três cenas prototípicas ou microcenários narrativos sobre as quais a unidade oracional se organiza.

Nós reencontramos aqui um agrupamento semelhante à distinção tradicional das classes acionais de Vendler (1967). O primeiro *cluster* aproxima-se do que Vendler denominou de *achievements*: eventos pontuais que expressam tipicamente uma mudança, mais ou menos súbita, de um estado para outro. As orações desse grupo, no *corpus*, ocorrem tipicamente associadas a sujeitos sintáticos não-agentivos e não intencionais. O segundo *cluster* agrupa o que Vendler denominou de Atividade e *Accomplishment*. Essas duas classes denotam processos que se desenvolvem no tempo, seja sem ou com um ponto télico (um ponto final ou de culminância do evento). Nos dados, eles ocorrem tipicamente associados com sujeitos sintáticos agentivos e intencionais. O *cluster* 3 denota o que Vendler chama de estado, o que equivale a uma eventualidade que se mantém inalterada em um determinado intervalo temporal.

A partir dessa divisão, é possível verificar a frequência relativa da sintaxe oracional em cada *cluster*. A Figura 8 mostra que cada grupo pode

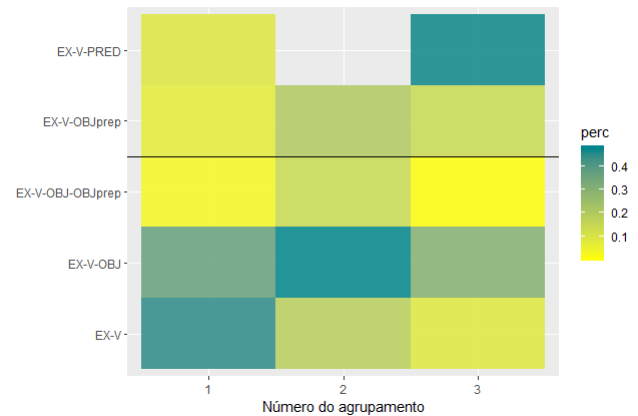


Figura 8: Frequência relativa da sintaxe oracional em cada *cluster*.

ser caracterizado pela predominância de determinadas formas sintáticas. Aqui a sintaxe é representada por um tipo de notação que agrupa unidades oracionais que compartilham certas características e comportamentos sintáticos semelhantes. Utilizou-se a notação “EXT” (argumento externo) como uma variável que representa o que é identificado classicamente como o sujeito sintático, independentemente da posição que ocupa na oração. Isso também inclui a desinência verbal, que em Português marca as noções gramaticais de sujeito, de pessoa e de número. O argumento externo pode representar também um sintagma fora do escopo da unidade oracional que tem um papel semântico associado ao verbo e à sua construção. O símbolo “V” representa uma unidade verbal, o que inclui não apenas o verbo, mas também perífrases aspectuais e modais, assim como construções compostas por verbos leves. Por fim, o símbolo “PRED” representa um sintagma predicativo, “OBJ” objeto direto não preposicionado e “OBJprep”, objeto indireto ou preposicionado.

O primeiro *cluster* (eventos pontuais que expressam uma mudança, mais ou menos súbita, de um estado para outro, tipicamente não agentivos e não intencionais) apresenta predominantemente estruturas do tipo “EX-V” como em (15) e (16). Mas podem ocorrer também formas “EX-V-OBJ”, como em (17), (18) e (19) especialmente envolvendo verbos de percepção (como ver e ouvir):

- (15) quatro pessoas morreu
- (16) depois a intimação estourô,
- (17) ela também viu ele.
- (18) Já ouvi passá uma sombra
- (19) Eu ganhei trinta mil reais

O segundo *cluster* (processos que se desenvolvem no tempo, com ou sem um ponto télico, tipicamente agentivos e intencionais) apresenta predominantemente estruturas do tipo “EX-V-OBJ” como em (20) e (21), mas ocorrem também com objetos preposicionados, como em (22) e (23). Em orações bitransitivas, como (23), o objeto preposicionado frequentemente marca o ponto télico do evento:

- (20) fiquei apertando esse ossinho
- (21) Aí eu preparei minhas mala toda.
- (22) eu tratava das criação
- (23) que ele me levô pro interior

Por fim, o terceiro *cluster* (eventualidade que se mantém inalterada em um determinado intervalo temporal, tipicamente não agentivas e não intencionais) é composto principalmente por orações com predicativos do sujeito como em (24) e (25), mas também com algumas ocorrências de estruturas do tipo “EX-V-OBJ”, principalmente com o verbo “ter”, como em (26) e certos verbos psicológicos como em (27):

- (24) Eu tô doida
- (25) E ele era evangélico,
- (26) eu tenho marido,
- (27) a psicóloga que sabe tudo,

5 Conclusão

Os resultados quantitativos apresentados nessa pesquisa mostram que as unidades oracionais, no português do Brasil, podem ser agrupadas em termos de parâmetros da transitividade, revelando a presença de três microcenários narrativos, semanticamente específicos, sobre os quais se desenrola o evento expresso. Apesar de não haver uma associação perfeita entre sintaxe e esses microcenários, é possível perceber a predominância relativa de certas estruturas sintáticas associadas a cada padrão semântico. Esse tipo de análise corrobora a hipótese adotada por diversos autores da Linguística Cognitiva (Brandt, 2004; Goldberg, 1995; Radden & Dirven, 2007) de que existe uma relação entre o núcleo conceitual de um determinado evento e a forma como ele é expresso em construções gramaticais.

Cada *cluster* analisado revela um tipo de significado protoconceptual, o que inclui traços aspectuais e actanciais próprios, que introduz as categorias lexicais da oração em uma cena ou cenário dinâmico. Essa noção de cenas predicativas, que vem desde Tesnière (1959), tem sido am-

plamente reconhecida no âmbito da Linguística Cognitiva.

Não existe um consenso na literatura, mesmo com o extenso debate produzido sobre o assunto, em relação a quais seriam essas cenas associadas à sintaxe oracional e como elas podem ser descritas em termos de valores semânticos. O desafio teórico é a demonstração de regras gerais das operações sintáticas, uma vez que os efeitos de significação que elas produzem são enormemente variados. Mas se tomarmos o caminho inverso, ao analisar a semântica de maneira relativamente independente da sintaxe, fica evidente que esses cenários micro-narrativos existem enquanto um grupo de certos traços associados. A questão central é que esses cenários aparecem correlacionados a certos padrões sintáticos, mas não são exclusivos destes últimos. Diferentes padrões sintáticos podem acomodar um mesmo padrão semântico geral, impondo a este último, possivelmente, particularidades.

A metodologia estatística de análise de agrupamentos mostrou-se uma ferramenta útil para se captarem esses padrões semânticos, chegando a resultados semelhantes às categorias aspectuais teoricamente conhecidas e mostrando, além disso, como essas categorias aspectuais se relacionam com categorias actanciais de agentividade e intencionalidade. Sob uma nova perspectiva, esse tipo de metodologia pode ser útil na investigação dos padrões semânticos a que os falantes são expostos, de maneira relativamente independente da sintaxe, e sugere uma arquitetura específica sobre a qual a língua se organiza.

Referências

- Bois, John W. Du. 2003. *Preferred argument structure: Grammar as architecture for function*. John Benjamins.
- Brandt, Per Aage. 2004. Dynamic schematism and the cognitive semantics of language. <https://case.edu/artsci/cogs/larcs/documents>, accessed on 19 March, 2019.
- Broccias, Cristiano. 2013. Cognitive grammar. Em Thomas Hoffmann e Graeme Trousdale (ed.), *The Oxford handbook of construction grammar*, 149–161. Oxford University Press.
- Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*, vol. 2. Cambridge University Press.
- Divjak, Dagmar & Nick Fieller. 2014. Cluster analysis: Finding structure in linguistic

- data. Em Dylan Glynn & Justyna Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 405–441. John Benjamins Publishing Company. doi 10.1075/hcp.43.16div.
- Givón, Talmy. 2001. *Syntax: An introduction*, vol. 1. John Benjamins Publishing.
- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
- Gower, John C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 857–871. doi 10.2307/2528823.
- Hair, Joseph F., William C. Black, Barry J. Babbin & Rolph E. Anderson. 2009. *Multivariate data analysis*. Prentice-Hall.
- Halkidi, Maria, Yannis Batistakis & Michalis Vazirgiannis. 2001. On clustering validation techniques. *Journal of intelligent information systems* 17(2-3). 107–145. doi 10.1023/A:1012801612483.
- Halliday, Michael Alexander Kirkwood, Christian Matthiessen & Michael Halliday. 2014. *An introduction to functional grammar*. Routledge.
- Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 251–299. doi 10.2307/413757.
- Johnson, Mark. 2013. *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.
- Lakoff, George. 2008. *Women, fire, and dangerous things*. University of Chicago Press.
- Lepesqueur, Marcus. 2017. *Transitividade na esquizofrenia: comparação dos relatos orais de eventos psicóticos entre grupos clínico e não clínico*: Universidade Federal de Minas Gerais. Tese de Doutorado.
- Lima, Lucia Chaves de Oliveira. 2013. *A transitividade na conversação: uma abordagem cognitivo-funcional*: Universidade Federal do Rio Grande do Norte. Tese de Mestrado.
- Lucena, Nedja Lima & Maria Angélica Furtado Cunha. 2012. Relações de herança em orações transitivas: O mecanismo de extensão metafórica. *Letras & Letras* 27(1).
- Mingoti, Sueli Aparecida. 2017. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. UFMG.
- Næss, Åshild. 2007. *Prototypical transitivity*, vol. 72. John Benjamins Publishing.
- R Development Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. <http://www.R-project.org>. ISBN 3-900051-07-0.
- Radden, Günter & René Dirven. 2007. *Cognitive English Grammar*, vol. 2. John Benjamins Publishing.
- Rousseeuw, Peter J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20. 53–65. doi 10.1016/0377-0427(87)90125-7.
- Rousseeuw, Peter J. & L. Kaufman. 1990. *Finding groups in data: An introduction to cluster analysis*. Wiley Online Library.
- Rozas, Victoria Vázquez. 2004. Transitividade prototípica y uso. *Boletín de Lingüística* 92–115.
- Shahrokhi, Mohsen & Ahmad Reza Lotfi. 2012. Manifestation of transitivity parameters in persian conversations: a comparative study. *Procedia-Social and Behavioral Sciences* 46. 635–642. doi 10.1016/j.sbspro.2012.05.176.
- Sokal, Robert R. & Peter H. A. Sneath. 1963. *Principles of numerical taxonomy*. W. H. Freeman.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Klincksieck.
- Thompson, Sandra A & Paul J. Hopper. 2001. Transitivity, clause structure, and argument structure: evidence from conversation. Em *Frequency and the emergence of linguistic structure*, vol. 45, 27–60. John Benjamins.
- Varela, Francisco, Evan Thompson & Eleanor Roch. 1991. *A mente corpórea: ciência cognitiva e experiência humana*. Instituto Piaget.
- Vendler, Zeno. 1967. *Linguistics in philosophy*. Cornell.