

TRANSFORMAÇÕES LEXICO-SEMÂNTICAS CORRELATAS À INFLUÊNCIA DA INTERNET

JESIEL SOARES SILVA (UFVJM)¹
LUIZ HENRIQUE MENDES BRANDÃO (UFMG)²

RESUMO: Neste trabalho, objetivou-se analisar as transformações ocorridas no uso da língua(gem) por parte de seus usuários tendo como base o período correspondente ao início dos anos 90, momento histórico em que a internet ainda não havia sido popularizada no mundo, em comparação ao ano de 2017, período marcado pelo amplo acesso à internet, principalmente nos países mais desenvolvidos. Para tal, realizou-se uma investigação tendo como base o COCA (*Corpus of Contemporary American English*) com o intuito de se verificar, através da associação de palavras com seus colocados (palavras justapostas a outras com uma frequência maior que o acaso) que, como alguns termos eram utilizados antes da popularização da internet e após o mesmo fenômeno. Através da análise estatística dos insumos, foi possível identificar que certos termos da língua (neste caso da língua inglesa) passaram a ser utilizados mais frequentemente para expressar algo relacionado à tecnologia, tendo sido os sentidos anteriores, bem como a frequência de uso de determinados termos, rebaixados, nesta transformação semântica, a uma frequência menor ou muito menor de uso após a realidade do acesso amplo à internet. Isso representa uma transformação léxico-semântica propiciada por um fenômeno de alcance global que influencia a vida das pessoas de modo a ressignificar o uso que fazem do mundo e conseqüentemente a metalinguagem que utilizam nas trocas que realizam com o mesmo.

PALAVRAS-CHAVE: Neologismos. Internet. Colocados.

ABSTRACT: *In this work, we aimed at analyzing the transformations which happened in the usage of the language by its users based in the period of time comprehending the beginning of the 90's, a historical moment when the internet was not popularized in the world, in comparison with the year 2017, year when people had already widespread access to the internet, mainly in the developed countries. To do so, we conducted an investigation based on COCA (Corpus of Contemporary American English) to find, through the association of some words with their collocates (words habitually juxtaposed with another with a frequency greater than chance), how some words were used before the popularization of the internet and after the same phenomenon. Through the statistical analysis of the input, it was possible to identify that certain words of the language (in this case, of the English language) started to be used more frequently to express something related to technology, having the previous meanings, as well as the frequency of certain words, been downtrended, with this semantic transformation, and now used with a lower or much lower frequency after the coming of the widespread access to the internet. This represents a lexicosemantic transformation made possible by a widespread phenomenon that influences people's lives resignifying the usage they make of the world and consequently the metalanguage they use in the interactions they have with it.*

KEYWORDS: *Neologisms. Internet. Collocates.*

1 Professor adjunto na Universidade Federal dos Vales do Jequitinhonha e Mucuri. E-mail: jesielsoaressilva@gmail.com

2 Mestrando em Psicolinguística pela Universidade Federal de Minas Gerais. E-mail: luizwhitemetal@yahoo.com.br

INTRODUÇÃO

A linguagem humana está em constante transformação. A veracidade desta afirmação pode ser verificada ao se refletir sobre o uso que fazemos da língua através de períodos de tempo não necessariamente muito longos. Palavras ganham novas representações semânticas e outros significados mais tradicionais podem recuar em frequência de utilização. Este processo cria o que em linguística chamamos de neologismos, que é a criação de novos termos derivados ou não de outros ou novos significados para termos já existentes. Segundo Castellvi (2006), há vários tipos de neologismos, como o formal, sintático e o que abordamos neste artigo, ou seja, o tipo semântico, que, segundo a autora, são gerados a partir de uma modificação de uma base léxica. Ainda segundo Castellvi (2006), os neologismos semânticos também podem surgir a partir de nomes próprios, usados como nomes comuns, como *Tupperware* ou *Armani*. A própria estrutura morfológica de termos em um sistema linguístico (neologismos formais) pode sofrer alterações por ocasião de reformas ou outras adaptações relevantes e ainda, dentre outras mudanças que podem ser diagnosticadas, percebemos que novos itens lexicais são criados ou importados de outras línguas sofrendo ou não adaptações (neologismos por empréstimo). Um exemplo disso pode ser observado na língua portuguesa, em palavras como *site*, *chat*, *web* e *internet*. Tais vocábulos são empréstimos da língua inglesa que se tornaram populares no português brasileiro contemporâneo. Tal fato pode ser indicio dos constantes avanços tecnológicos que têm abraçado a maior parte da civilização humana, principalmente nas últimas décadas. As novas possibilidades advindas no bojo destes avanços tecnológicos têm se mostrado um dos mais potentes motores de transformação linguística atualmente, sobretudo após o início da popularização da internet a partir da década de 1990 (MODIS, 2005). Além desta variável, há muitas outras que contribuem para sua dinamicidade.

Além da incorporação de novos vocábulos a uma língua, também é possível associar a influência da internet e sua metalinguagem a uma transformação semântica em determinados vocábulos que passam a expressar significados novos e mais abrangentes. Podem-se destacar palavras como *baixar* e *conectar*, que, após o advento da rede mundial de computadores, passaram a portar, além de sua carga semântica tradicional, outras novas como “Transferir para um computador = DESCARREGAR” e “Fazer uma ligação a um computador ou dispositivo ou a uma rede de computadores ou dispositivos”, respectivamente (DICIONÁRIO PRIBERAM DA LÍNGUA PORTUGUESA, 2008-2020).

Tanto para uma verificação mais fidedigna do fenômeno quanto para sua melhor compreensão, faz-se necessário lançar mão de uma investigação empírica da língua em uso através de diferentes períodos históricos, utilizando, para tal, uma fonte de dados com reconhecida confiabilidade e robustez. A partir da década de 1960, especificamente desde os empreendimentos de Sinclair (1966), a Linguística de *Corpus*, através de seu crescente número de colaboradores, tem se prestado a investigar a linguagem de forma notadamente empírica, criando compilações de *corpora* orais, escritas ou multimodais de diversos gêneros textuais capazes de responder satisfatoriamente à demanda supracitada.

Um exemplo deste tipo de ferramenta é o *Corpus Of Contemporary American English* (COCA) que, segundo Davies (2010), é o primeiro de tamanho grande e não estático (que recebe novos textos periodicamente) já criado e que, ao contrário de seus antecessores e muitos sucessores, pode ser utilizado para verificar mudanças linguísticas desde 1990 até a atualidade. Segundo Davies (2010), uma das mais importantes características do COCA que está ausente em outros *corpora* de renome é seu criterioso trabalho diacrônico por meio de inserções uniformes em todos os gêneros para cada período de tempo que cobre.

Segundo Kader e Richter (2013)³, através da linguística de *corpus*, e em especial em *corpora* como o COCA, somos capazes de manusear grandes quantidades de dados linguísticos de forma acelerada. Pode-se, assim, identificar palavras mais frequentes em um determinado período histórico e para um gênero textual específico, comparando essa frequência dentro de cada um destes filtros e em relação a outros vocábulos relevantes, o que pode levar a uma melhor compreensão de como esses termos foram sendo transformados ao longo de períodos de tempo e a um diagnóstico dos possíveis motores dessas transformações ao se comparar essas mudanças com aquelas ocorridas no mundo em determinados períodos históricos. Kader e Richter (2013) ainda citam a possibilidade de se descobrir padrões comuns de linguagem produzidos por aprendizes de uma segunda língua através de *corpora* de aprendizes como o *International Corpus of Learner English* (ICLE), que conta com inserções oriundas de aprendizes a nível avançado de língua inglesa falantes nativos de diversas línguas.

Há ainda outros *corpora*, como o *Longman learners corpus*, que contempla mais de 10 milhões de textos produzidos por aprendizes de língua inglesa de todos os níveis de proficiência. O estudo com *corpus* se revela altamente eficiente na aprendizagem e ensino de línguas, uma vez que pode apresentar aos aprendizes os contextos naturais de ocorrência das palavras, sintagmas, frases ou expressões da língua, explicitando, assim, diferenças de significado entre palavras, significados de uma palavra, como são usadas e sua frequência, para que se saiba se é uma palavra que caiu em desuso ou não, dentre outras vantagens.

Levando em conta estes aspectos, neste artigo, buscamos entender as mudanças na carga semântica de alguns vocábulos no que tange à frequência de seus colocados dentro do *corpus*. A nossa hipótese é de que a diferença na frequência dos colocados de 1990 em relação aos vocábulos-alvo é significativa com relação a 2017, sugerindo uma interferência da intervenção da evolução tecnológica no uso destes vocábulos em diversos gêneros de texto.

JUSTIFICATIVA/TEORIA

Conforme descrito por Sardinha (2004), a linguística de *corpus* lida com investigações linguísticas por meio de evidências empíricas dela obtidas em compilados de dados linguísticos. Esses bancos de dados provêm insumos da língua em contextos reais de seu uso. Starvik (1996) apresenta a seguinte citação de Leech a respeito do assunto:

a linguística de *corpus* não define somente uma metodologia emergente para o estudo da linguagem, mas uma nova maneira de fazer pesquisa, e de fato uma nova abordagem filosófica para este assunto. O computador, como uma ferramenta tecnológica de poder indiscutível, tornou este novo tipo de linguística possível (LEECH, 1992, p. 106 *apud* STARVIK, 1996, p. 12).

Alinhando-se a essas proposições aqui apresentadas, pode-se entender a linguística de *corpus* como algo que vai além de uma simples ferramenta de pesquisa e, conforme a fala de Leech, estamos diante de uma nova abordagem filosófica para o assunto. De fato, a partir principalmente da publicação da obra *Beginning the study of lexis*, de Sinclair (1996), percebe-se uma guinada nos estudos linguísticos para uma perspectiva mais empírica. Segundo Oliveira

³ KADER, Carla Callegaro Corrêa; RICHTER, Marcos Gustavo. *A Coesão semântica entre as lexis relacionadas aos fatores de atribuição em textos acadêmicos da área de letras e psicologia*. Acervo da UFSM. Santa Maria, 2013. 26 p. Disponível em: <<http://jararaca.ufsm.br/websites/l&c/download/Artigos13/04Carla.pdf>> Acesso em 13 nov. 2019. Não publicado.

(2009), os estudos de *corpora* são caracterizados pela busca de tendências, probabilidades ou padrões de ocorrência ao lidarem com grande quantidade de dados provenientes desses sistemas computadorizados, porém os estudos de *corpus*, segundo a autora, não devem ser concebidos como meros resultados numéricos e estatísticas extraídos do *corpus*, pois a tarefa de interpretação dos dados muitas vezes deve levar em conta também aspectos socioculturais que estão ligados ao texto. Oliveira (2009) ainda propõe que a linguística de *corpus* deve ser considerada uma teoria exofórica, o que significa que não está isolada em seu universo de estudo, mas estabelece interface com outros, como o da linguística aplicada, cujos estudiosos podem se valer do avanço nos estudos com *corpora* para determinar métodos mais eficazes de intervenção pedagógica. Psicolinguistas podem estudar determinados tipos de comportamentos dos falantes através de observações em *corpora*, dentre outras trocas.

O avanço da tecnologia, segundo Oliveira (2009), é um dos principais fatores que têm contribuído para o desenvolvimento do campo da linguística de *corpus*. O que era feito nos primórdios do desenvolvimento do campo com fichas com conteúdo linguístico guardadas em caixas, conforme Svartvik (1996), passou hoje a um sofisticado sistema de arquivamento e análise por meio eletrônico em *corpora* com tamanhos que ultrapassam 100 milhões de trechos linguísticos e que são utilizados amplamente para diversos fins como ensino/aprendizagem de língua estrangeira, estudos tradutórios, dentre inúmeras outras possibilidades. Atualmente, as compilações de *corpora* têm crescido no mundo todo, contando, além de *corpora* da língua geral de vários idiomas, também com exemplares especializados, como o *News On the Web (NOW)*, que traz textos apenas de revistas e jornais online, o *TV corpus*, que contém texto apenas de episódios de programas de TV e o *The Wikipedia corpus*, que contém o texto completo do site *Wikipedia*. Estes três exemplos citados estão entre os *corpora online* mais acessados do mundo.

Uma característica da maioria dos estudos de *corpora* citada por Oliveira (2009) é de serem feitos tomando como base o léxico, o que, segundo ela, se justificaria pela oferta maior de programas que auxiliam nesse tipo de análise. Entretanto, nos últimos anos, vários programas que fazem a marcação gramatical automática de um *corpus* foram desenvolvidos e disponibilizados, fator que facilita análises mais completas como a da função sintática das palavras (*parsing*) e etiquetagem (*tagging*). McEnery *et al* (2019) citam ferramentas como o Lancsbox e o Sketch engine. Através de qualquer uma dessas duas ferramentas os linguistas, ainda que sem treinamento em programação, podem agrupar palavras com suas variantes, produzir textos através de um sistema de reconhecimento de fala que consegue distinguir palavras mesmo que produzidas de forma truncada ou incompleta ou se valer de sofisticados mecanismos de análise estatística imbutidos em programas como os citados acima. Tem-se, então, um universo de estudos amplo e crescente.

Francis e Kucera (1982) e Biber (1988) entre outros, assim como Oliveira (2009) citam a dinamicidade no desenvolvimento de teorizações e pesquisas dentro da linguística de *corpus* desde seu surgimento. Sanchez (1995) propõe uma resposta sobre o que é linguística de *corpus* que descreve a linha de pesquisa como se segue:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador com a finalidade de propiciar resultados vários e úteis para descrição e análise (SANCHEZ, 1995, p. 8-9).

A história da compilação de *corpus*, apesar de sua expressão moderna ser bem recente, já existia através do *corpus* helenístico, compilado na Grécia antiga sob Alexandre, o grande e das citações bíblicas que eram compiladas na idade média. Berber Sardinha (2010) atesta tal informação. A realidade contemporânea da análise de *corpora*, porém, lida com processos computacionais eletrônicos. Berber Sardinha (2004) faz menção ao *Taggit*, que foi um sistema etiquetador morfossintático computacional apresentado no começo da década de 1970 que praticamente iniciou essa nova era de estudos. A partir deste que é considerado um divisor de águas, cada vez mais têm se desenvolvido as ferramentas disponíveis para os pesquisadores com sistemas eletrônicos cada vez mais sofisticados.

Ainda a partir dos estudos de Berber Sardinha (2004), pode-se desprender um fato importante sobre as pesquisas científicas realizadas com *corpora* que diz respeito a sua maior ocorrência através de análise de palavras que estão associadas diretamente ao termo, o que se denomina *colocados*. Segundo Berber (2004) uma palavra deve ser julgada por sua companhia e isso é uma das razões para a ampla utilização deste tipo de investigação tão útil para se aferir a recorrência de palavras. Pode-se medir, através de ferramentas estatísticas, a associação de termos que se deseja entender, bem como se essa ligação se faz por aleatoriedade ou não.

Seguindo a linha pioneira de desenvolvimento da área da linguística de *corpus*, encontramos Kjellmer (1994) e Benson e colaboradores (1986), que elaboraram respectivamente o *A dictionary of English collocations: based on the Brown Corpus* (KJELLMER, 1994), um dicionário exclusivo de colocados que teve *corpus* como base e o *BBI combinatory dictionary of English* (dicionário combinatório BBI de inglês), que é um dicionário de combinações.

O COCA tem sido base para inúmeras investigações científicas desde sua criação. Para citar artigos recentes feitos com base nesta ferramenta online, elencamos o trabalho de Davies e Kim (2019) que consistiu em estudos sobre mudanças ocorridas ao longo do tempo em construções chamadas em inglês de *INTO-CAUSATIVE constructions* e a pesquisa de Robinson e Duncan (2019) para averiguar variações sintáticas em perguntas denominadas em língua inglesa de *Wh-all questions*.

Uma pesquisa recente empreendida por Davies e Kim (2019) utilizou o COCA em conjunto com outro *corpus* online para averiguar uma série de mudanças de ordem lexical, semântica e sintática em construções causativas na língua inglesa denominadas de *INTO-CAUSATIVE constructions*, que são, basicamente, construções causativas com a palavra *into*. Ainda se investigou se o referido construto se tornou mais ou menos comum com o passar do tempo com base nos dados de *corpora* e através de análises estatísticas. Levando em conta essas concepções e seguindo as definições de Partington (1998, p. 16-17), nosso trabalho se encaixa no estudo colocacional de ordem estatística, que averigua a probabilidade significativa de ocorrência de alguns colocados em relação a um termo chave.

METODOLOGIA

A análise da presente pesquisa foi feita com dados extraídos do *Corpus Of Contemporary American English* (COCA). Esta ferramenta é um *corpus* online e representa, segundo o próprio site (www.english-corpora.org),

O *Corpus of Contemporary American English* (COCA) é o único *corpus* de inglês americano grande balanceado por gênero. O COCA é provavelmente o *corpus* de inglês mais utilizado e está relacionado a muitos outros *corpora* de inglês que

criamos que oferecem *insight* sobre a variação na língua inglesa sem paralelos (DAVIES, 2008-2020).⁴

A base do COCA, ainda segundo o próprio site, contém mais de 560 milhões de palavras de texto, sendo 20 milhões por ano de 1990 até a atualidade, que é o período de escopo do *corpus* até a data de publicação deste artigo, porém, trata-se, segundo Davies (2010) de um *corpus* dinâmico, o que significa que, de forma diversa daqueles estáticos, o COCA é alimentado periodicamente com novos textos, o que significa que podemos prever que estará sempre atualizado.

Os gêneros nos quais o site está igualmente dividido são textos: *spoken* (falados), *fiction* (de ficção), *popular* (populares), *magazines* (de revistas), *newspapers* (de jornais) e *academic* (acadêmicos).

O site disponibiliza buscas por frequências de palavras isoladas, expressões inteiras e *collocates* (colocados), que são palavras que ocorrem juntamente a uma palavra chave, ainda há a possibilidade de se comparar a ocorrência de palavras lado a lado e se determinar através de filtro as palavras que ocorrem com uma palavra chave em determinada posição de proximidade a ela. Ainda pode-se filtrar gênero e ano juntos ou separadamente, dentre outras formas de busca possíveis. Em todas estas buscas, sempre há a apresentação do contexto em que cada uma das palavras está ocorrendo nos textos originais fontes da base de dados, o que colabora para a otimização da compreensão da língua alvo.

Para analisar os dados de forma quantitativa, utilizamos testes paramétricos (assumindo a normalidade dos dados, devido à natureza específica dos mesmos). Composto nosso conjunto de testes aplicados no conjunto de dados, fizemos uso da média aritmética, juntamente com o desvio padrão das amostras. Para averiguar a significância estatística das diferenças entre amostras, utilizamos o *test-t*, que, segundo Kremelberg (2010) é um método, talvez o mais utilizado, para avaliar a diferença entre as médias de dois grupos. O *test-t* foi aplicado em dois contextos distintos, primeiramente em relação à frequência e, em seguida, em relação ao MI score. Por fim, para entender a relação entre os grupos de amostra e gerar os *insights* acerca dos resultados obtidos, fizemos o uso da correlação de Pearson (*r*), que, segundo Kremelberg (2010), tem o propósito de determinar se há uma relação significativa (correlação) entre duas variáveis contínuas. Ele representa uma correlação que vai de -1 a 1. -1 significa uma correlação negativa perfeita entre duas variáveis (Se uma aumenta a outra diminui), 0 (zero) quer dizer que não há uma dependência linear de uma para com a outra e 1 representa a correlação positiva perfeita entre elas (se uma aumenta, a outra também).

ANÁLISE DE DADOS

Como previamente descrito, nosso primeiro procedimento de análise foi o cálculo da média paramétrica (*M*) dos três primeiros colocados mais frequentes dos vocábulos-alvo, o cálculo feito considerou dois aspectos: i) quanto à época (tanto o ano de 1990 quanto 2017); ii) quanto à especificidade (a ocorrência geral desses colocados no *corpus*, bem como a ocorrência deles em função do vocábulo-alvo). Em seguida, aplicamos um teste paramétrico (*test-t*) para verificar se as diferenças de médias em relação às duas épocas são significativamente diferentes.

⁴ Traduzido de: *The Corpus of Contemporary American English (COCA) is the only large, genre-balanced corpus of American English. COCA is probably the most widely-used corpus of English, and it is related to many other corpora of English that we have created, which offer unparalleled insight into variation in English.*

Abaixo segue uma tabela com dados expositivos sobre quais foram os três colocados mais frequentes dos nossos vocábulos-alvo concernente aos anos de 1990 e 2017 respectivamente:

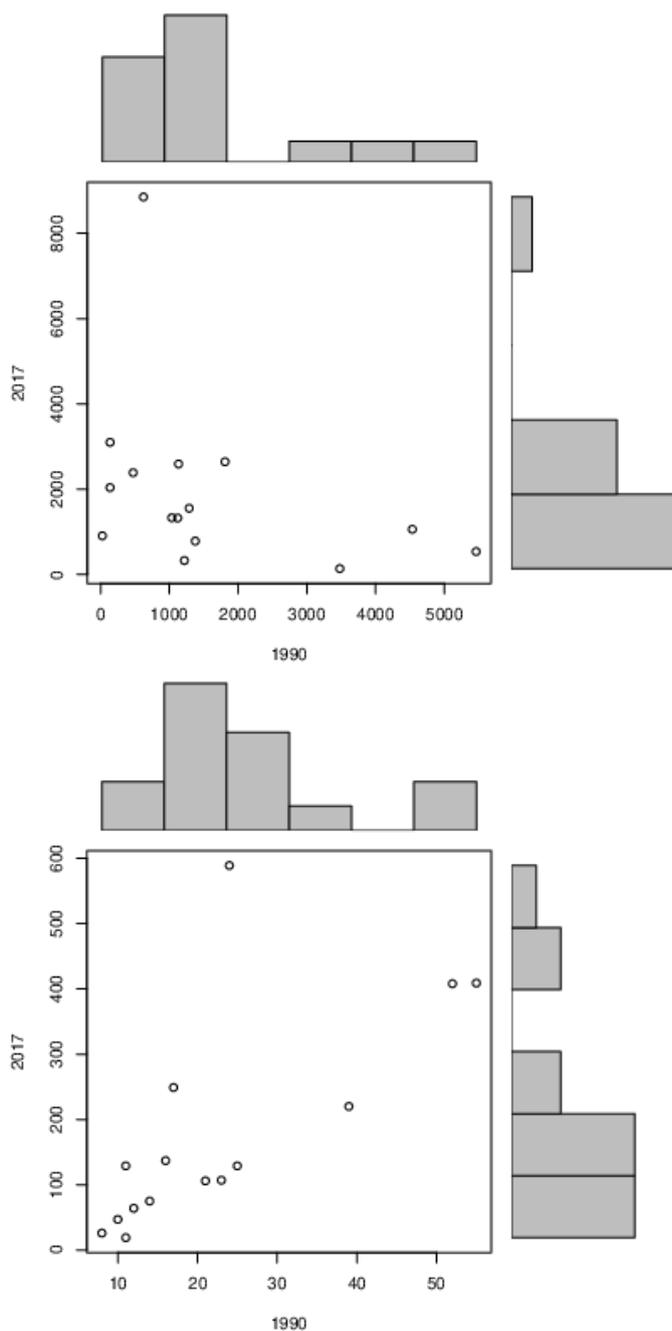
SITE		WEB		ACCESS		CLOUD		FEED		
	freq		freq		freq		freq		freq	
1990	construction	25	spider	11	equal	55	dust	24	store	14
	near	23	life	8	information	52	aort	17	food	12
	waste	21	intricate	11	gain	39	smoke	16	hungry	10
2017	web	129	site	129	digital	409	web	5896	Facebook	75
	browse	107	page	26	unlimited	408	NPR.ORG	249	post	64
	uses	106	services	19	internet	220	CHECK	137	Twitter	47

Tabela 1: Disposição dos colocados dos vocábulos-alvo

Para obtermos uma visão geral do comportamento dos nossos vocábulos-alvo com seus colocados, performamos um *test-t*, considerando as médias gerais dos colocados em relação aos vocábulos-alvo nas duas épocas distintas (1990-2017). Corroborando com o que foi previsto na apresentação das hipóteses, a) a média das palavras tidas como colocados, mas analisadas de forma isolada no *corpus*, considerando as duas épocas distintas, não apresentariam um efeito específico, pois não se configurariam como duas categorias distintas e b) O *test-t* revelou: $t(14)=0,28$, $p>0,05$. Como podemos observar no p-valor, não há diferenças estatisticamente importantes entre estes dois grupos (ou seja, p-valor maior do que .05). Partimos, então, para testar nossa segunda hipótese, de que, quando analisadas como colocados dos nossos vocábulos-alvo, em relação às duas épocas distintas, estes dois grupos apresentariam diferença significativa. O *test-t* revelou: $t(14)=3,91$, $p<0,005$. Como pudemos observar no p-valor, há um efeito entre estes dois grupos, mostrando um indício da transformação das ocorrências desses colocados em função das palavras com uma carga semântica relacionada à tecnologia. Esse dado é extremamente representativo daquilo que propomos neste artigo. É um dado que tanto esclarece quanto evidencia, nesse dado específico, que as palavras realmente têm uma transformação diacrônica em relação à evolução da tecnologia e que fatores externos interferem na evolução da língua. Isso é um indício de que transformações histórico-sociais têm interferência direta nas transformações linguísticas.

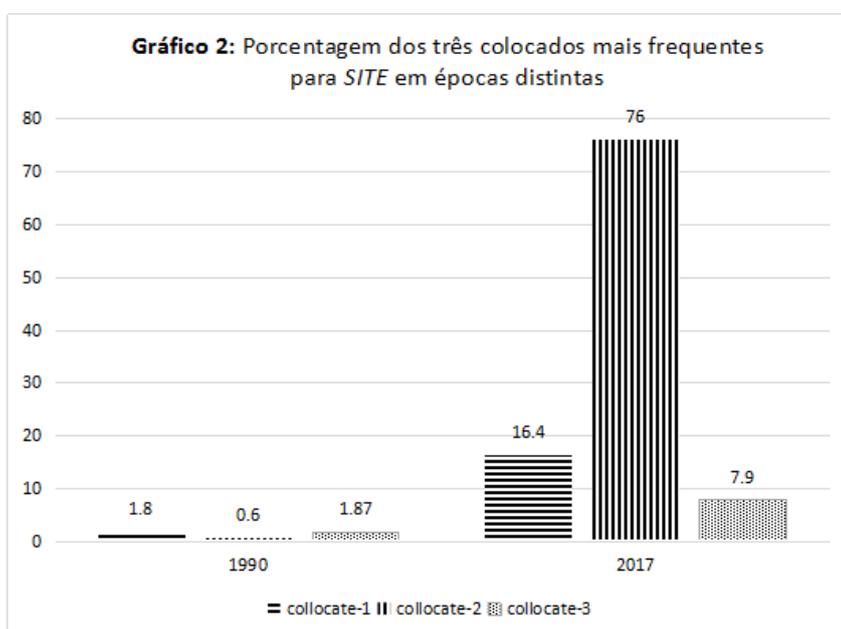
Para confirmarmos os resultados previamente apresentados, realizamos um teste paramétrico de correlação utilizando o coeficiente *r* de Pearson, aplicados nas mesmas situações das análises acima. O resultado para os dados gerais mostrou $r=-0.3617$ e os dados como colocados mostraram $r=0.6822$. Dessa forma, fica atestado que: a) esse resultado observado anteriormente, sobre as palavras alvo passarem a estar mais associadas a colocados ligados à tecnologia, é compartilhado para todas elas, ou seja, não é um fenômeno aleatório e b) os primeiros colocados associados a cada uma das palavras alvo em 2017 são altamente frequentes, enquanto os mesmos eram mais dispersos no começo da década de 1990. Abaixo, dispomos o gráfico de dispersão para exemplificar as correlações propostas.

Gráfico 1a: Correlação dos colocados como vocábulos isolados e **Gráfico 1b:** juntamente com os vocábulos-alvo



Partindo deste primeiro insight dos resultados e do efeito encontrados, fizemos a análise das médias dos três primeiros colocados mais frequentes para os nossos vocábulos-alvo considerando a diferença entre os anos de 1990 e 2017. Aplicamos o *test-t* (teste T) para averiguar a significância destes dois grupos de média representando as duas diferentes épocas para cada vocábulo-alvo separadamente. O *test-t*, conforme mencionamos no campo sobre a metodologia, foi aplicado em dois contextos distintos, primeiramente em relação à frequência e, em seguida, em relação ao MI score. Apresentaremos os dados a seguir, bem como a disposição visual das porcentagens, para uma compreensão maior dos resultados.

Em relação ao vocábulo-alvo *site*, o resultado do *test-t* pareado para frequência de ocorrência dos três colocados mais frequentes (*construction, near, waste*) em relação ao ano de 1990 (M=26,3, SD=4,1) e os três colocados mais frequentes em relação ao ano de 2017 (*web, browser, uses*) (M=114, SD=9,6), foi $t(2)=11,1$, $p<0,01$. Já em relação ao MI score para a mesma amostra, em 1990 (M=4,7, SD=0,8) e em 2017 (M=8,1, SD=1,6) foi $t(2)=2,31$, $p<0,05$. Esse grupo de dados mostra de forma evidente, clara e estatística, que há um efeito de separação entre os dois grupos, ou seja, em relação à frequência dos colocados, se constituem em dois grupos diferentes. Os dados ainda mostram que os colocados sobre tecnologia em 2017 são muito mais evidentes e específicos. Abaixo, o gráfico representativo da porcentagem em relação ao mesmo conjunto de dados:

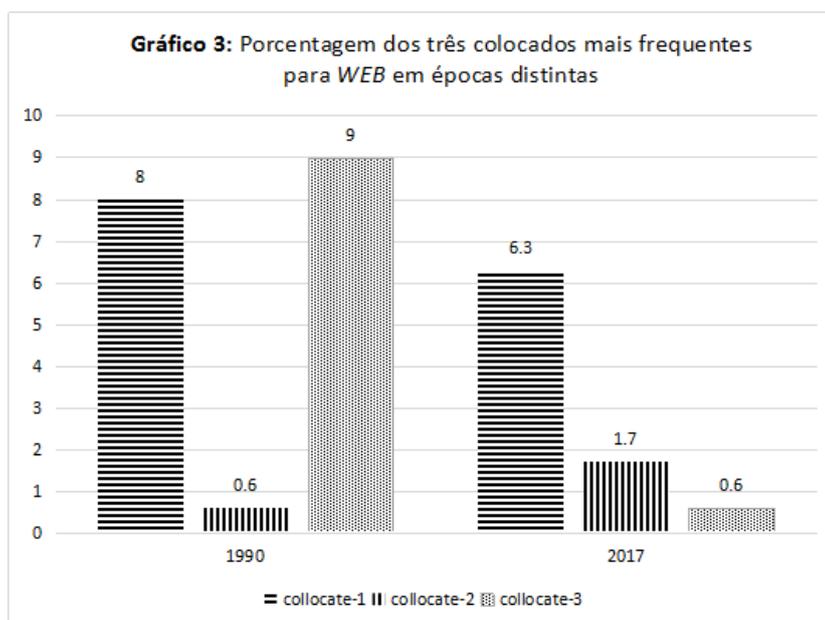


Percebemos, no gráfico 2, que a frequência dos três colocados mais frequentes no ano de 2017 ocorrem de forma notadamente superior aos de 1990, o que mostra um indício de uma aderência maior desses colocados com o vocábulo *site*. Considerando que os três colocados mais frequentes para o vocábulo *site*, em 1990, foram respectivamente *construction, near, waste*, e que estes colocados não necessariamente remetem a uma semântica em torno de conceitos relacionados à tecnologia, percebemos que há um espraiamento dessa frequência, não apresentando um efeito expressivo.

Já quando notamos os três colocados mais frequentes ao vocábulo *site* em 2017 (*web, browser, uses*), pelo menos os dois primeiros deles são evidentemente relacionados à tecnologia; a porcentagem de frequência deles é notadamente maior e percebemos esse fator de aderência junto ao vocábulo. Isso mostra que, ao longo do tempo, esses colocados ocorrem com mais frequência junto ao vocábulo-alvo.

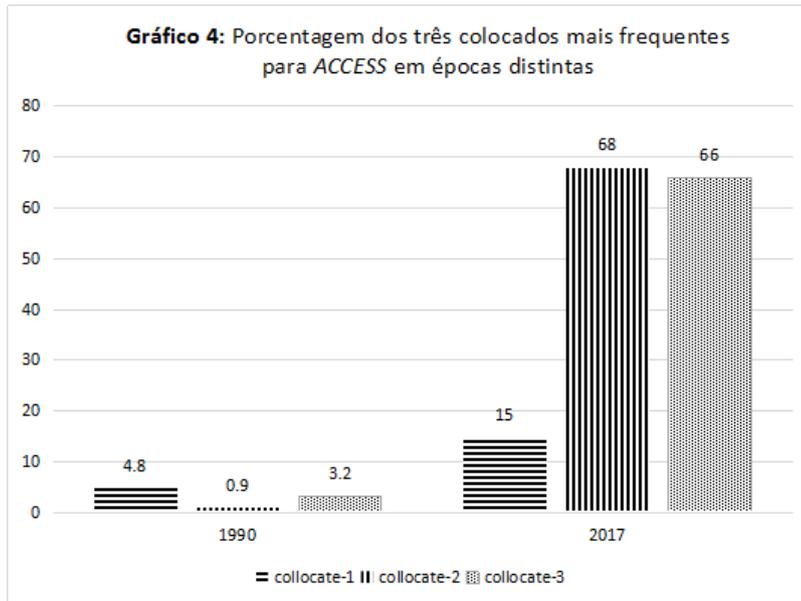
Em relação ao vocábulo-alvo *web*, o resultado do *test-t* pareado para frequência de ocorrência dos três colocados mais frequentes (*spider, life, intricate*) em relação ao ano de 1990 (M=10,3, SD=1,7) e os três colocados mais frequentes de 2017 (*site, page, services*) (M=58,5, SD=27,4), foi $t(2)=1,36$, $p<0,05$. Como podemos observar, a diferença apresentada no teste paramétrico mostrou-se estatisticamente significativa. Tal resultado é representativo da nossa hipótese inicial, ou seja, as palavras que selecionamos para análise neste artigo passaram a ser

acompanhadas de colocados relacionados à tecnologia após a popularização da internet e isso de uma forma muito mais frequente do que as associações anteriores a este período, que foram rebaixadas em sua frequência de uso em conjunto com esses termos. Já em relação ao MI score para a mesma amostra, em 1990 (M=8,1, SD=3,6) e em 2017 (M=5,9, SD=1,2) foi $t(2)=2,31$, $p<0,05$. Da mesma forma, quando considerado o MI score, a diferença se manteve significativa, o que também é um indício dessa regularidade apresentada sobre a especificidade dos colocados em relação às palavras. Abaixo está o gráfico 3, que é representativo da porcentagem em relação ao mesmo conjunto de dados



O gráfico 3 mostra um cenário um pouco diferente em relação à aderência dos colocados ao vocábulo *web*, dois dos três colocados (*spider e life*) apresentaram maior adesão, proporcionalmente, ao vocábulo-alvo do que os três primeiros colocados de 2017 (*site, page, services*). Isso pode ser um indício da própria especificidade da palavra, que possui colocados mais comumente ligados à ela, independente da época ou da evolução tecnológica. Outra possível explicação é a de que o vocábulo *web* já traz consigo uma significação completa, não necessitando de colocados para complementar o significado.

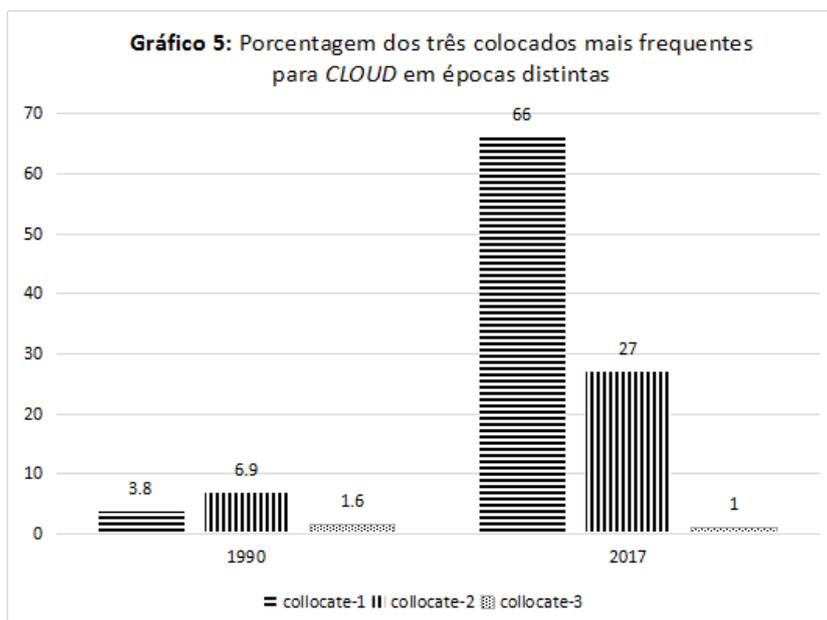
Em relação ao vocábulo-alvo *access*, o resultado do *test-t* pareado para frequência de ocorrência dos três colocados mais frequentes (*equal, information, gain*) em relação ao ano de 1990 (M=18,5, SD=2,7) e no ano de 2017 (*digital, unlimited, internet*) (M=2094, SD=3281), foi $t(2)=1,09$, $p<0,005$. Como podemos ver, a diferença apresentada é significativa de forma evidente com um p-valor abaixo de 0,005. Já em relação ao MI score para a mesma amostra, em 1990 (M=8,6, SD=3,2) e em 2017 (M=5,1, SD=0,8) foi $t(2)=0,36$, $p<0,05$. Da mesma forma, o MI score apresentou significância em relação à diferença entre 1990 e 2017. Abaixo, o gráfico representativo da porcentagem em relação ao mesmo conjunto de dados:



Assim como no gráfico 2, o gráfico 4 mostra a grande diferença em relação à adesão dos três primeiros colocados em relação ao vocábulo *access* concernente a 1990 e a 2017. Em 2017, os colocados *digital* e *internet* apresentam evidências notáveis de aderência ao vocábulo *access*, confirmando nossa hipótese acerca da frequência dos vocábulos com carga semântica relacionada à tecnologia.

Em relação ao vocábulo-alvo *cloud*, o resultado do *test-t* pareado para frequência de ocorrência dos três colocados mais frequentes (*dust*, *oort*, *smoke*) em relação ao ano de 1990 (M=48,5, SD=8,3), e os três colocados mais frequentes de 2017 (*web*, *check*, *storage*) (M=345,2, SD=108), foi $t(2)=5,12$, $p<0,005$. Este resultado evidencia também uma diferença significativa entre os dois grupos de colocados.

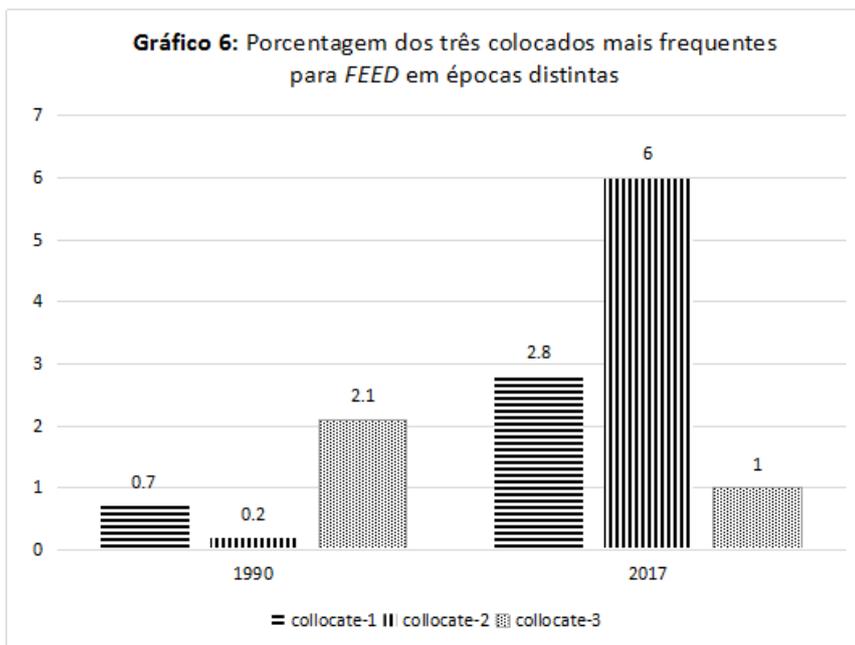
Já em relação ao MI score para a mesma amostra, em 1990 (M=8,8, SD=2,8) e em 2017 (M=7,3, SD=2,1) foi $t(2)=0,87$, $p>0,05$. Abaixo, o gráfico representativo da porcentagem em relação ao mesmo conjunto de dados:



Assim como nos gráficos 2 e 4, o gráfico 5 mostra a evidente diferença no que concerne à adesão dos três primeiros colocados em relação ao vocábulo *cloud* concernente a 1990 e a 2017. Em 2017, os colocados *web* e *storage* apresentam medidas evidentes de aderência ao vocábulo *access*, o que também confirma nossa hipótese sobre a frequência dos vocábulos com carga semântica relacionada à tecnologia.

Em relação ao vocábulo-alvo *feed*, o resultado do *test-t* pareado para frequência de ocorrência dos três colocados mais frequentes (*store, food, hungry*) em relação ao ano de 1990 ($M=12,2, SD=2,1$) e os mais frequentes de 2017 (*facebook, post, twitter*) ($M=62,3, SD=14$), foi $t(2)=7,14, p<0,05$. Como podemos observar, os dados são representativos da nossa hipótese e foi possível constatar que no ano de 2017 os colocados das palavras alvo estão relacionados à tecnologia e ainda estão mais agrupados.

Já em relação ao MI score para a mesma amostra, em 1990 ($M=4,9, SD=1,4$) e em 2017 ($M=7,9, SD=0,8$) foi $t(2)=1,55, p>0,05$. Neste conjunto de dados, apenas a diferença das ocorrências se mostrou significativa, enquanto a diferença no MI score não mostrou significância estatística. Abaixo, o gráfico representativo da porcentagem em relação ao mesmo conjunto de dados.



Assim como nos gráficos 2, 3 e 5, o gráfico 6 evidencia a diferença na adesão dos três primeiros colocados em relação ao vocábulo *feed* concernente a 1990 e a 2017. Em 2017, os colocados *Facebook* e *Twitter* apresentam medidas evidentes de aderência ao vocábulo *access*, o que também confirma nossa hipótese sobre a frequência dos vocábulos com carga semântica relacionada à tecnologia, de que após a internet ser popularizada, termos antes associados a atividades outras comuns no dia a dia passaram a ser fortemente associados à tecnologia.

CONCLUSÃO

Os dados, de forma geral, confirmaram nossa hipótese sobre a diferença na frequência dos colocados relacionados à tecnologia concernente aos anos de 1990 e 2017. Quanto ao gênero (*speaking, academic* etc), não foi possível fazer uma análise de forma separada porque esta opção apenas estava disponível na versão paga do COCA e estávamos utilizando apenas

a versão livre, entretanto os dados que analisamos apontam que há muito a ser verificado neste mesmo conjunto de dados se os gêneros forem separados e analisados de forma autônoma.

Conclui-se, ainda, a partir dos colocados, que as transformações observadas na utilização dos vocábulos alvo desta investigação também se dão no campo semântico, pois as palavras que mais comumente acompanhavam esses termos alvo em 1990, ou seja, antes da democratização do acesso à internet, cedem lugar, atualmente, em questão de frequência, a novas companhias, palavras que foram adjungidas a esses vocábulos no advento da era da internet e que se tornaram mais frequentemente associadas aos termos-alvo do que os colocados de outrora. No ano de 2017, as pessoas devem associar termos como *site* e *web* mais a uma página na internet e ao próprio ambiente virtual do que a um local físico ou uma teia de aranha, por exemplo.

Conforme exposto nos gráficos de número 2, 4 e 5, houve ainda um aumento na própria frequência das palavras alvo. Futuras investigações podem ser feitas para se apurar mudanças ocorridas com outros termos ou possivelmente até com expressões linguísticas e outros construtos linguísticos. O mesmo pode ser feito para averiguar mudanças hipoteticamente causadas por outros fenômenos ocorridos no mundo.

Tais contribuições possibilitam uma maior compreensão não só dos processos de mudança pelos quais os termos de uma língua passam de acordo com as experiências humanas no mundo, como também, por consequência, da extensão da influência histórica de determinados fenômenos nas mais diversas áreas do fazer humano.

REFERÊNCIAS

- BENSON, M., BENSON, E., ILSON, R. (orgs.) **The BBI dictionary of english word combinations**. Amsterdã/Filadélfia: John Benjamins, 1986.
- BIBER, D. **Variation across speech and writing**. Cambridge: Cambridge University Press, 1988
- Davies, Mark. **The Corpus of Contemporary American English (COCA): 600 million words, 1990-present, 2008**. Disponível em: <https://www.english-corpora.org/coca/>. Acesso em: 19 fev. 2020.
- CASTELLVI, Maria Teresa CABRÉ. **La clasificación de neologismos**. Alfa, São Paulo, 50 (2): 229-250, 2006
- DAVIES, Mark. The Corpus of Contemporary American English as the first reliable monitor corpus of English. **Literary and Linguistic Computing**, Brigham, v. 25, n. 4, 2010. Disponível em: <https://academic.oup.com/dsh/article-abstract/25/4/447/997323?redirectedFrom=fulltext>>. Acesso em: 21 ago. 2019.
- FRANCIS, W. N.; KUCERA, H. **Frequency analysis of English usage: lexicon and grammar**. Boston: Houghton Mifflin, 1982
- DAVIES, Mark; KIM, Jong-Bok. Historical shifts with the INTO-CAUSATIVE construction in American English. **The Gruyter mouton**, [S.L.], v. 57, n. 1, 2019. Disponível em: <http://web.khu.ac.kr/~jongbok/research/2019/2019-ahci-into-historical-shift-linguistics.pdf>> Acesso em 21 ago. 2019
- DICIONÁRIO PRIBERAM DA LÍNGUA PORTUGUESA**. Desenvolvido por Lello editores, Porto, 1996 e 1999. Licenciado à Priberam em 2008. Disponível em: < <https://dicionario.priberam.org/sobre.aspx>> Acesso em 21 ago. 2019
- KJELLMER, G. A. **A dictionary of English collocations: based on the Brown Corpus**, v. 3. Oxford: Oxford University Press, 1994
- KREMELBERG, David. **Practical statistic: a quick and easy guide to IBM © SPSS © Statistics, STATA, and other statistical software**. Sage: Los Angeles, 2011.
- MC ENERY, Tony, et al. **Corpus Linguistics, Learner Corpora, and SLA: Employing Technology to Analyze Language Use**. **Annual Review of Applied Linguistics** (2019), 39, 74–92
- MODIS, Theodore. The end of the internet rush. **Technological Forecasting & Social Change**, Lugano, v. 72, n. 8, 2005. Disponível em: < <https://www.sciencedirect.com/science/article/pii/S0040162505000843>> Acesso em: 21 ago. 2019
- OLIVEIRA, Lúcia Pacheco de. Linguística de corpus: Teoria, interfaces e aplicações. **Matraga**, Rio de Janeiro, v. 16, n. 24, 2009. Disponível em: < <https://www.e-publicacoes.uerj.br/index.php/matraga/article/view/27796>>. Acesso em: 21 ago. 2019
- PARTINGTON, A. **Patterns and meanings: using corpora for english language research and teaching**. Amsterdã/Filadélfia: John Benjamins, 1998

- ROBINSON, Mary; DUNCAN, Daniel (2019) Holistic Approaches to Syntactic Variation: Wh-all Questions in English. **University of Pennsylvania Working Papers in Linguistics**: v. 25, n. 1, 2019. Disponível em: <<https://repository.upenn.edu/pwpl/vol25/iss1/23/>>. Acesso em: 21 ago. 2019
- SANCHEZ, A. Definición e historia de los corpus. In: SANCHEZ, A. et al. (orgs.). **CUMBRE**: corpus lingüístico de español contemporáneo. Madri: SGEL, 1995, p. 7-24.
- BERBER SARDINHA. T. **Linguística de Corpus**. Barueri, SP: Manole, 2004.
- SINCLAIR, J. McH. **Beginning the study of lexis**. In: BAZELL, C. E. In memory of R. Firth. Londres: Longman, 1966, p. 410-430.
- SVARTVIK, Jan. Corpora are becoming mainstream. In: THOMAS, J. and SHORT, M. (orgs.). **Using corpora for language research**. London and New York: Longman, 1996. p 3-13.

Recebido em 16-11-2019.

Aceito em 12-02-2020.