

Relação entre produtividade e frequência na construção do significado

Mário Eduardo Viaro¹

¹Departamento de Letras Clássicas e Vernáculas – Universidade de São Paulo (USP)
Av. Luciano Gualberto 403 – 05508-900 – São Paulo – SP – Brasil

maeviaro@usp.br

Abstract. *This paper describes the variation, which resulted from several interpretations of 4627 virtually possible Portuguese monosyllabic strings. The theoretical distinction between meaning and sense is also examined, because it does not allow one to analyze that problem comprehensively.*

Keywords. *Monosyllables; Semantics; Phonology; meaning; sense.*

Resumo. *Este artigo descreve a variação resultante de várias interpretações de 4627 seqüências monossilábicas virtualmente possíveis em português. Examina-se a distinção teórica entre significado e sentido, uma vez que ela não permite fazer uma análise exaustiva desse problema.*

Palavras-chave. *Monossílabos; Semântica; Fonologia; significado; sentido.*

1. Introdução

Quando trabalhamos com Semântica, costumamos afirmar que algumas palavras *têm* significado, enquanto outras *adquirem* um sentido pelo contexto. Essa divisão é questionável, pois dizer que uma palavra *tem* significado é imaginá-la como um recipiente (LAKOFF & JOHNSON 1980). Dessa forma, *pefiz* não tem significado, mas, numa frase como *peopequê peeu pefiz*, decodifica-se um sentido por meio de uma regra que diz “toda sílaba começa por *pê*”. No entanto, significado e sentido não são em essência distintos. O conceito de sentido como um significado *lato sensu* faz-nos pensar que para uma palavra ter significado são precisos critérios, que nem sempre são fáceis de determinar. O mais usual é o de que a palavra seja amplamente utilizada e conhecida por todos os falantes da língua, mas a falta de uma visão onisciente impede-nos de falar disso seriamente. A razão dessa bipartição reside, na verdade, num fundo epistemológico e tradicional e o idealismo platônico, modificado em requintadas formas de modelos religiosos e filosóficos, passou tacitamente aos modelos científicos, inclusive à Lingüística. Do ponto de vista do receptor, ao menos, seria mais tranqüilo utilizar um único termo, o da *interpretabilidade*, presente tanto nas palavras que dependem, quanto nas que independem de um contexto. Quando surgiram os modelos estruturalistas na Fonologia, entendeu-se o fonema de forma ideal, face à variação ilimitada dos sons ou fones, tentando, dessa forma, reproduzir algo paralelo ao que ocorre com as capacidades inatas da linguagem, pois, na confecção de signos, abstraem-se inúmeros elementos não-significativos em busca de uma certa constância (TROUBETZKOY 1970). Essa constância, porém, é artificial quando a essa abstração inata somam-se outras, culturais, advindas de inúmeras convenções arbitrárias. Discordamos, portanto, que haja um ponto de partida na língua associado a uma

variante neutra ou livre de fatores sociolingüísticos, históricos e discursivos. Esse pressuposto nem sempre está muito claro, mas muitas vezes subentende-se nas teorias.

2. Metodologia

Querendo conhecer a produtividade do sistema fonológico português e quantos monossílabos há nessa língua, criamos um programa em linguagem Java, que gerasse apenas seqüências monossilábicas virtualmente possíveis no português, seguindo os seguintes critérios: (a) O primeiro elemento do *onset*: *p, t, k, b, d, g, f, v, s, z, ʃ, ʒ, m, n, R, l* ou \emptyset ; (b) O elemento líquido do *onset*: *r* ou *l*, que só se junta com *p, t, k, b, d, g, f*, com uma restrição de não gerar **tl, *dl*; (c) Uma semivogal no *onset*: *w, j* ou \emptyset , que não se associa a determinadas vogais, evitando *ji, wo, wɔ* e *wu*; (d) Uma vogal: *a, e, ε, i, o, ɔ* e *u*. Convém observar que as seqüências geradas *ti, di* podem ser interpretadas como [tʃi] e [dʒi]; (e) Um elemento nasal na coda: *N* ou \emptyset , gerando as vogais nasais, exceto **eN* ou **ɔN*; (f) Uma semivogal na coda: *w, j* ou \emptyset , evitando *ij, ɔw* e *uw*. As formas geradas como *-ɔl, -ul* podem ser interpretadas como [ɔw], [uw]. Também não se geraram **eNw, *iNw, oNw* (possível interpretação de *oN*), *eNj* (interpretação de *eN*), *uNj*; (g) Um elemento líquido na coda: *l, r* ou \emptyset , que não se associa a determinadas vogais, evitando assim vogal+**Nl, *Nr, *wl, *wr, *jl* e **jr*. Não se geram também **el* e **ol*; (h) Um elemento sibilante na coda, representado por *s* (interpretáveis como [s], [z], [ʃ] ou [ʒ]) ou \emptyset , evitando **rs* e **ls*. Algumas palavras especiais também foram geradas (*lhe, lhes, nhô, nha, gol, gols* etc.). O resultado se mostrou um pouco deficitário no caso dos monossílabos não-*standard* (cf. abaixo) por não gerar encontros como *wo, wou, uNj*, mas esse detalhe só foi percebido *a posteriori*; mesmo assim, as conclusões deste artigo continuam válidas. Ao todo, formaram-se 4627 seqüências. Chamemos esse universo de *total de seqüências monossilábicas*. Se eliminarmos a semivogal do *onset*, o número de palavras cai para 1872 (ou 40,46% do total), que chamaremos *total de seqüências monossilábicas puras*. Essa distinção é importante, uma vez que os monossílabos impuros são muitas vezes interpretáveis, por diérese, como dissílabos: *'pjaNw* ou *pi'aNw*. A averiguação de quais palavras têm significado foi submetida a 156 pessoas treinadas na leitura do alfabeto fonético, a saber, alunos de Graduação da disciplina Fonética e Fonologia no ano de 2004 e sobre esse resultado é que se fundamentam as análises subseqüentes.

3. Monossílabos *standard*

Há algumas palavras que inegavelmente são reconhecidas como monossílabos, exceto por algum lapso de memória ou distração, fato muito comum diante de uma lista de palavras descontextualizadas, problemas contra os quais nem mesmo o mais perspicaz lingüista está imune. Por isso, foi preciso estabelecer um critério pautado em fatores assumidamente artificiais. Diante de uma variantes, um monossílabo foi arbitrariamente escolhido como *standard* e outro não, ex.: *dar* e *da* são significantes para o mesmo infinitivo do verbo “dar”. Optou-se, nesse caso, pelas formas com *-r* na coda como *standard*. Nos casos dos monossílabos impuros, como *pjãw*, optou-se pela interpretação “pião” como *standard* e “peão” como não-*standard*, porque a segunda tem mais representações fonéticas que a primeira. Idem *kwajs* tem interpretação *standard* “quais” e não-*standard* “coais”. Se trabalharmos apenas com monossílabos puros, esses últimos casos desaparecem. Assim, *pas* tem dois significados *standard*: “pás” e “paz”,

mas *pajs* tem, ao lado do *standard* “pais” também os não-*standard* “paz” e “pás”. Parece razoável que flexões da mesma palavra estejam juntas na mesma categoria. A finalidade desse critério foi registrar ao menos uma vez a existência do monossílabo, uma vez que as formas não-*standard* são de difícil tratamento, como será visto. Em nossa lista, obtivemos 524 monossílabos *standard*, que correspondem a 11,32% do total de seqüências monossilábicas. Se não levarmos em conta a semivogal do *onset*, o número abaixa para 416 monossílabos *standard* puros, que é um valor percentualmente maior (22,22%) com relação ao total de seqüências monossilábicas puras. É, portanto, possível afirmar que, em português, a pureza da seqüência monossilábica é diretamente proporcional à capacidade de gerar monossílabos. Essa regra é válida também para os monossílabos não-*standard*.

Uma crítica poderia ser feita ao fato de termos optado pela inclusão do *l* como elemento da coda, opondo-se a *w*. Fato é que algumas variantes opõem *l:w* e outras só têm *w* e essa diferença fundamenta praticamente dois subsistemas dentro do português. Para evitar a solução dos subsistemas, que seria indesejável (por razões igualmente pouco ligadas à Lingüística, mas a um conceito de língua una, vigente desde o nacionalismo do século XIX), prefere-se falar de neutralização (na perspectiva estruturalista) ou de uma transformação de vocalização (da perspectiva gerativista). O inventário de fonemas é, assim, normalmente visto dogmaticamente como fixo. Também é fato que outras variantes opõem *s* pré-dorso-alveolar (grafado *ç*) do *s* ápico-alveolar (grafado *s* ou *ss*). Por serem desprestigiadas, limitando-se a algumas aldeias no norte de Portugal, nunca se afirma que as variantes com um único *s* (esmagadora maioria de falantes lusófonos) seja uma neutralização dos dois *s* (embora historicamente isso faça sentido). Como se vê há aí razões de vária ordem (sobretudo política) na manutenção de um único sistema para uma língua, do ponto de vista sincrônico, porém isso infelizmente não condiz com os fatos. Com referência a nosso *corpus*, se desejamos trabalhar com seqüências sem *l*, que só aceitem *w*, é preciso também gerar *ɔw* e *uw*: ao todo seriam 4393 seqüências monossilábicas (1725 puras), dos quais 518 são *standard* (11,79%) e 410 puros (23,77%). A neutralização aumenta, por meio desses dados, o valor da interpretabilidade; isso, porém, é ilusório.

4. Monossílabos raros e freqüentes

Para que um monossílabo seja de fato modelar para diversos fenômenos (sobretudo a analogia) é preciso associá-lo à sua freqüência. Assim, *piar* é uma seqüência impura que está na lista dos monossílabos *standard*, de modo que *piei*, *piou*, *piais* e *piai* também entram. Também aí estão *pá* e seu plural *pás*. Se quiséssemos levantar apenas as formas lematizadas, deveríamos deixar de lado toda flexão e trabalhar apenas com as formas invariáveis, as no singular e os verbos no infinitivo, ou seja, 231 monossílabos ou 5% de toda a lista (ou 195 puros, 10,42%), números úteis somente para quem quisesse fazer um glossário mínimo dos monossílabos portugueses. De fato, se introduzirmos o fator *freqüência* veremos que algumas formas *standard* têm uma interpretabilidade variável: em vários testes, observamos que poucos falantes do português reconhecem *piai* como um monossílabo português. A freqüência de uso de *piai* como “segunda pessoa do plural do imperativo do verbo *piar*” é baixíssima. O mesmo se pode pensar de *ruais*, que, embora a gramática normativa aponte como defectivo, é passível de uma interpretação como “segunda pessoa do plural do presente do subjuntivo do verbo *ruir*”. No entanto, por coerência do método, ambos são

standard. O fato de ser *standard* não implica, portanto, que seja um monossílabo freqüente e facilmente interpretável. Há também formas não-*standard* que são freqüentes.

Também é preciso apontar que os monossílabos *standard* foram obtidos pelo teste e não por meio de um dicionário. Qualquer léxico cita palavras como *cris*, espécie de punhal malaio, não-*standard* e raro, apesar de aceito por critérios normativos (da mesma forma que *ruais* não é aceito). Há listas colossais de monossílabos, como a de CASANOVAS (1954), que podem entrar no cômputo de outras pesquisas, diferentes da que é apresentada neste artigo. Em nossos monossílabos, trabalhou-se com 197 formas no singular (ou 37,6% do total de *standard*), 145 no plural (27,67%), 23 infinitivos (4,39%), 162 formas verbais flexionadas (30,92%) e 54 formas invariáveis (10,31%). Se falarmos de formas puras, esses números são: 175 no singular (42,07%), 130 no plural (31,25%), 10 infinitivos (2,4%), 105 formas verbais flexionadas (25,24%) e 53 formas invariáveis (12,74%). A soma ultrapassa 100% porque há muitos casos de dupla e até tripla interpretação da mesma seqüência, como *pus* que pode ser um substantivo singular ou uma forma verbal; ou *næs*, que pode ser, pelo menos, o pronome *nós*, o plural de *nó* ou o substantivo singular *noz*. Dizemos “pelo menos”, porque somente esses elementos morfológicos podem ser levados em conta, para fins quantitativos, uma vez que averiguações dos casos de homonímia e polissemia tornariam nosso trabalho inviável. Assim *piw* tem dois valores *standard*: tanto a primeira pessoa do presente do indicativo do verbo *piar* quanto um nome singular, seja ele o deverbal do verbo *piar* (“o pio do passarinho”) quanto um adjetivo (“o fiel era muito pio”).

5. Monossílabos não-*standard*

Pesquisando no buscador de internet *Google*, é possível observar que há muitas pessoas com sobrenome *Piai* e também se trata de um topônimo. Esses fatos dependem de um conhecimento compartilhado que não é extensivo a todos os interpretantes. Também essas interpretações não decorrem da aplicação de flexões morfológicas. Trata-se do grupo chamado de monossílabos não-*standard* aos quais se acrescentam as variantes: se *pes* é *standard*, então *pejs* foi computado como variante. Dessa forma, temos ao todo 1968 monossílabos não-*standard*, que equivalem a 42,53% das seqüências monossilábicas. Foram detectados 1193 monossílabos não-*standard* puros, ou 63,72% das seqüências monossilábicas puras. A soma de monossílabos *standard* e não-*standard* perfaz o número de 2185, ou 47,22% do total das seqüências monossilábicas. Desses, 307 (6,63%) podem ter interpretações *standard* ou não-*standard*, 217 (4,69%) só têm interpretações *standard* e 1661 (35,9%) só têm interpretações não-*standard*. Das seqüências geradas, 2442 (52,78%) não tiveram nenhum tipo de interpretação. Se pensamos em seqüências puras, as porcentagens são mais altas: a soma dos monossílabos *standard* e não-*standard* sobe para 1609 (85,95%), dos quais 260 (13,89%) têm interpretação *standard* e não-*standard*, 156 (8,33%) são exclusivamente *standard*, 933 (49,84%) são exclusivamente não-*standard* e 523 (27,94%) não têm nenhuma interpretação.

Deduzimos, a partir do exposto, que a pureza é inversamente proporcional à interpretabilidade de uma seqüência monossilábica. Dito de outra forma, as seqüências puras têm visivelmente menos possibilidade de não ter nenhum significado. Isso se justifica pelo fato de apenas *kw* e *gw* serem *onsets* em que a semivogal tem o mesmo *status* de *r* e *l*, ao passo que outros casos, como *sw*, *tw*, *kj*, *pj* podem ser interpretados

como *su*, *tu*, *ki*, *pi* gerando dissílabos oxítonos. Nosso *corpus* dispõe de 39 seqüências com *kw-*, das quais 36 têm algum significado, sendo apenas 3 *standard* (*qual*, *quais*, *quão*). Das 39 *gw-* apenas 5 têm algum significado, não sendo nenhum *standard*. Desproporções como essas impediram-nos de descartar totalmente as formas impuras.

Uma seqüência como *sis* pode, assim, ter inúmeras interpretações (à parte das polissemias ou homonímias possíveis e de valores não constantes na pesquisa como nome de uma cidade da Turquia). Em uma frase como “*este compasso tem dois sis e três dós*”, teremos uma interpretação *standard*, mas é possível encontrá-la em: [sis’kara’i] *esses cara aí*, [sis’forver’dadzɨ] *se isso for verdade*, [eli’fojpra’sis] *ele foi pra Assis*, [telefo’nejpra’sis] *telefonei pra Ciça*. É sempre cômodo encontrar aí aféreses e apócopes a partir de uma fala pausada, platonicamente ideal, no entanto, não foi esse o caminho que trilhamos. Nesses casos todos, não-*standard*, estamos diante da mesma seqüência, ainda que não o sejam suprasegmentalmente. Fato é que essas realizações existem, pois são flagradas em diversos momentos comunicativos, tanto na interpretação de falsas análises, quanto na criação de chistes ou em métrica. A mesma seqüência foi apontada como uma sigla, com variada homonímia (Serviço de Informação de Segurança, Serviço Integrado de Supervisão etc.) ou como um elemento de formação (prefixo *cis-* em oposição a *trans-*). A categoria dos monossílabos não-*standard* é, pois, bastante ampla e passível de uma subclassificação.

Nenhum nome próprio foi incluído entre os *standard* e na sua mesma categoria entram apocorísticos. Alguns são muito comuns, como *zɛ* para José, *dri* para Adriana ou *pri* para Priscila, outros menos comuns como *bjɛ* para Gabriel, *zu* para Zuleide ou Zulmira, outros ainda bastante raros como *tɔl* para Marisol etc. Vale dizer que entre nomes próprios entraram também marcas de produtos, bandas de música, cidades etc. Avaliar qual deles é *standard* seria muito arbitrário. Foram apontados 526 monossílabos nessa categoria (26,73% dos monossílabos não-*standard*) ou 420 puros (35,21%). Os nomes próprios têm comportamento especial, sendo possíveis até algumas seqüências não previstas, como *vla* para Vladimir. Apenas um dado ilustraria suas peculiaridades: o *onset* *z* revela-se muito pouco produtivo, sendo apenas 6 formas *standard*, todas com vogal *e* ou *o*. No caso das formas não-*standard* há 50, com todas as vogais, das quais metade são nomes próprios. Da mesma forma, nenhum estrangeirismo entrou no cômputo das formas *standard*, pois se uns são comuns como *gej*, outros são menos comuns como *piN*. Nenhuma palavra estrangeira foi aceita simplesmente por ser reconhecida: é preciso ter uso no português. Assim, um informante bilíngüe de catalão reconheceu *pɛw* como “pê”, mas não se aceitou, devido à consciência externada de que a palavra pertencia a outro sistema. Foram apontados 238 monossílabos nessa categoria (12,09%) ou 190 puros (15,93%). A questão da consciência tanto na etimologia do estrangeirismo quanto no truncamento dos apocorísticos fez-nos optar por incluir essas categorias como não-*standard*. O mesmo se pode dizer das siglas, que representam 66 casos (3,35%) ou 58 puros (4,86%). Um informante, diante de apocorísticos, siglas e estrangeiros não só está apto a dizer que significam, mas também de onde vêm. O mesmo se pode dizer de expressões como *estou pê da vida*, *ele tá fu*. Algumas seqüências como *bi* têm vasta homonímia (=bilhão, bissexual, bicha etc.). Há certo titubeio em afirmar que essas seqüências têm significado, embora ninguém lhe negue um sentido, de modo que se pode afirmar que *o sentido está diretamente ligado à consciência etimológica da palavra*. Quando não sabemos a etimologia da palavra, diz-se que sem dúvida *tem* um

significado. Curiosíssimo é o caso das seqüências de palavras como [sis] *se isso*. É muito difícil determinar quantas são, mas nosso *corpus* aponta para 497 (25,25%). Trata-se do único caso em que as seqüências puras apontam para um percentual menor, pois só 149 foram encontradas (12,49%). De fato, *di, kuN, si, pru, i, u, us* se tornam facilmente por sinérese *dj, kw, sj, pw, j, w, ws*, formando monossílabos impuros: *djuN* “de um”, *kwa* “com a”, *sj* “se”, *pwaw* “pro Al(fredo)”, *jew* “e eu”, *war* “o ar”, *ɛws* “é os” etc. Também nesse caso há consciência da aglutinação e a escrita (ou o letramento) tem grande papel nessa consciência.

Um caso especial são as onomatopéias: 198 (10,06%) ou 150 puras (12,57%). Como os nomes próprios, muitas fogem até mesmo do sistema fonológico do português e não puderam ser previstas. Há conseqüências curiosas em sua formação: a consoante *l* aparece na segunda posição do *onset* em apenas 14 palavras *standard*, mas em 224 não-*standard*, das quais 31 são onomatopéias.

6. Conclusões

Cumpra observar que há ainda uma gama enorme de variantes que entram nos monossílabos não-*standard*, mas muitos foram controladas. Nos totais, não se incluiu nenhum exemplo de lambdacismo: os 1113 casos com *r* na 2ª posição do *onset* poderiam converter-se em outros tantos casos de monossílabos com *l*. Idem para o rotacismo tanto nos 795 casos de *l* no *onset* ou para os 350 casos de *l* na coda. Bastaria, para tal, imaginar um informante que consistentemente transformasse $-l > -r$ e isso de fato só existe nas caracterizações de personagens ou em textos humorísticos. Muitos apontaram como variantes de *krwɛl*, as formas *krwɛw*, *krwɛr* e *krwɛ*, com seus respectivos plurais. É pouco provável - embora seja perfeitamente possível imaginar - que alguém pronuncie a palavra erudita “dual” como [ˈdwaɹ], mas isso está na base de muitas argumentações simplistas a respeito da realização dos fonemas. É válido afirmar que há uma grande projeção sobre a fala do outro e se essas transformações fossem tão simples, os metaplasmos seriam altamente consistentes diacronicamente e não dependeriam de fatores externos como o fato de a origem da palavra ser culta ou não. Também termos jocosos e errados segundo a norma culta resultam num certa variação na interpretabilidade. Indagados alguns informantes sobre a seqüência equivalente a *pãos*, alguns diziam que “essa forma não existe” (não tem significado?), outros diziam que “muita gente fala assim”, que “conheço pessoas que falam assim” ou que “é possível que alguém fale assim”. Outro caso curioso: a seqüência *pãe*, reconhecida por apenas 6 informantes como “misto de pai e mãe”, embora dotada de sentido, não era unanimemente vista como portadora de significado. Como visto acima, isso decorre não só pelo fato de ser um termo jocoso (e, portanto, ligado ao discurso e não ao sistema), mas também devido à consciência etimológica de sua formação. São fatores subjetivos como esses que sustentam, a nosso ver, a distinção entre sentido e significado.

7. Referências bibliográficas

- CASANOVAS, C. F. de Freitas. *Pequena enciclopédia de monossílabos*. 2.ed. Rio de Janeiro: São José, 1954.
- LAKOFF, G.; JOHNSON, M. *Metaphors we live by*. Chicago: The Chicago University Press, 1980.
- TROUBETZKOY, Nikolai. *Principes de phonologie*. Paris: Klincksieck, 1970.