
UM ESTUDO ESTATÍSTICO SOBRE AS OXÍTONAS NO PORTUGUÊS

ANA LÍVIA DOS SANTOS AGOSTINHO*

GABRIEL ANTUNES DE ARAÚJO**

RESUMO

O objetivo deste artigo é apresentar um estudo estatístico sobre as palavras com acento final (ou oxítonas) no português. Baseado em um *corpus* de 10.494 palavras (ou seja, todas as palavras nominais oxítonas) do *Dicionário Houaiss* e suas respectivas transcrições fonéticas, mostraremos que a maioria das palavras oxítonas do português é formada por empréstimos lexicais. Ainda que o latim, o “tupi”, o francês, o árabe e o iorubá tenham sido as principais fontes dos empréstimos, o português pegou emprestadas palavras de mais de cem línguas diferentes. Também discutiremos a qualidade do elemento final da oxítona (se vogal, glide ou consoante) e a frequência geral destas palavras.

PALAVRAS-CHAVE: oxítonas, português, fonologia, acento final.

INTRODUÇÃO

O objetivo deste texto é analisar as palavras oxítonas em português a partir de um levantamento estatístico. Partindo da hipótese segundo a qual as palavras com acento final foram incorporadas, sobretudo, via empréstimos, uma vez que não havia palavras oxítonas no latim, procuraremos determinar a origem dessas palavras (ano e língua), bem como analisar o elemento final e a sua frequência. O texto foi organizado da seguinte forma: na próxima seção, apresentamos a metodologia e a

* Professora da Universidade de São Paulo. A autora contou com financiamento da Fapesp, processo 2006/03038-5, a cuja instituição agradece.

E-mail: ana.agostinho@yahoo.com

** Professor da Universidade de São Paulo e Kyoto University of Foreign Studies (KUFS). O autor agradece ao Departamento de Estudos Luso-Brasileiros da KUFS pelo auxílio financeiro concedido a este trabalho.

E-mail: g.antunes@usp.br

formação do *corpus* e destacamos a diferença entre a lista-expandida e a lista-base. Na seção 3, são apresentadas e discutidas as análises estatísticas dos dados, considerando-se as línguas emprestadoras e as datas de entrada das palavras oxítonas no português. Nesse ponto, são consideradas em detalhe as dez línguas que mais contribuíram com empréstimos. Em seguida, apresentamos uma análise dos elementos finais das oxítonas e também uma discussão a respeito da frequência-web. A seção final traz algumas conclusões sobre este estudo e levanta questões para pesquisas futuras.

Devido à complexidade que envolve a formação de um *corpus* confiável, optamos por empregar um *corpus* consolidado formado por todas as palavras com acento lexical final registradas no *Dicionário Houaiss* (HOUAISS & VILLAR, 2001). Embora saibamos que haja palavras oxítonas que não estão dicionarizadas e, ao mesmo tempo, palavras dicionarizadas de origem literária ou pertencente à nomenclatura técnico-científica, e, por isso, de uso raro, preferimos manter o *corpus* dicionarizado por se tratar de um conjunto consistente de dados e estar disponível a todos os interessados, além de o *Dicionário Houaiss* (DH) ser o mais recente dicionário etimológico da língua portuguesa. Por mais raras ou excepcionais que sejam, as palavras do dicionário pertencem a um determinado registro da linguagem e, portanto, são válidas para o propósito de analisar a ocorrência das palavras com acento final no português. Dada a natureza deste trabalho, a informação etimológica do DH também nos é fundamental. Uma vez que, conforme Ferreira Netto (2001), não havia palavras oxítonas no latim, atribui-se a presença de palavras portuguesas oxítonas de origem latina a alterações fonológicas. Mas o que dizer das palavras de outras origens? Para tentar responder a essa questão, abordaremos os seguintes tópicos:

- a) A origem das palavras oxítonas, procurando verificar a porcentagem de termos latinos e de outras línguas em geral;
- b) A data de entrada no português e qual a língua emprestadora;

- c) A qualidade do elemento final (vogal, consoante ou glide);
- d. A freqüência das oxítonas.

A motivação para este trabalho surgiu a partir de uma passagem de um trabalho inédito de Sandalo (1999, p. 1), no qual a autora sugere que “o grande número de oxítonas [no português] não se pode explicar como sendo casos de empréstimos em sua maioria”. No entanto, como tentaremos mostrar, é possível afirmar que cerca de 80% das palavras oxítonas nominais do português são empréstimos e que, portanto, é possível, de fato, explicá-las como casos de empréstimos em sua maioria (AGOSTINHO, 2007).

O termo cedido por empréstimo é introduzido na língua alvo por falantes nativos que têm acesso à língua *emprestadora* na sua forma oral ou escrita. A nativização, ou adaptação dos empréstimos, é regida por padrões fonológicos da língua receptora. Ou seja, empréstimo é uma palavra simples ou composta (ou uma sentença) oriunda de uma língua emprestadora, incorporada ao discurso da língua receptora (adaptado de PARADIS e LABEL, 1994).

METODOLOGIA

O *corpus* desta pesquisa foi trabalhado com o auxílio da ferramenta computacional MatLab que nos proporcionou o *desmanche* da lista de palavras original do programa do DH.¹ Cabe aqui ressaltar que empregamos a noção lingüística de acento, e não a definição ortográfica. Diante das possibilidades que o programa nos permitia, foi-nos possível elaborar uma lista com todas as palavras oxítonas presentes no DH, a que denominaremos lista-expandida. Todo o trabalho de classificação dos verbetes e sua posterior separação (definindo o *corpus*), a natureza da sílaba final, qualidade das vogais e das consoantes finais, transcrição fonética e listagem da língua que deu origem ao termo português, no caso dos empréstimos, foi efetuado manualmente.

A lista-expandida² contém 37.591 palavras oxítonas. No caso de múltiplas acepções, o critério estabelecido foi o de considerar cada eventual entrada como múltipla, seguindo o critério do DH. Da lista com 37.591 palavras foram excluídos todos os verbos e as palavras formadas por composição ou derivação, o que resultou na lista-base. No DH, em razão de a forma verbal aparecer no infinitivo, todos os verbos são oxítonos, *a priori*, sendo, portanto, excluídos. Também foram excluídas da lista-base as palavras oxítonas que, ortograficamente, possuem hífen, como, por exemplo, *arte-menor*, uma vez que o hífen, como um possível sinal gráfico de composição, acarretaria a repetição da palavra oxítona.

As palavras evidentemente derivadas de outra palavra existente no português também foram excluídas. No entanto, há casos como o da palavra *aragonês*, proveniente do espanhol *aragonés*, na qual é possível identificar a presença do sufixo derivacional para gentílicos, ainda que a palavra *aragão* não exista no DH como forma simples. Portanto, qualquer palavra cuja forma no português seja adaptada diretamente da língua de origem, será mantida na lista final, chamada doravante de lista-base. Caso a palavra seja derivada no português e sua forma de base também esteja dicionarizada, o termo derivado não será computado como oxítono. Algumas palavras que são ortograficamente oxítonas, mas que possuem consoantes proibidas na posição de coda no português, também foram eliminadas, como, por exemplo *farad*, realizada foneticamente em grande parte dos dialetos do Português do Brasil (PB) como [fa.ˈra.dʒɪ], sendo, portanto, uma pseudo-oxítona. Excluindo-se todos os casos supramencionados, a lista-base foi formada por 10.494 palavras, contendo, assim, todas as palavras oxítonas que não são verbos, os substantivos compostos ou derivados (exceto os casos excepcionais mencionados) e as pseudo-oxítonas.

A lista contendo 37.591 palavras com acento final foi formatada no programa Microsoft Excel. Foram elaboradas colunas que nos permitiram verificar vários aspectos que cobriam os objetivos iniciais

do projeto. A lista-base contém mais de 200 páginas, enquanto que a lista-expandida possui cerca de mil páginas.

A lista-base é composta apenas pelas palavras consideradas aptas a partir da análise da décima primeira coluna da lista-expandida, o que quer dizer que inclui todas as palavras que não sejam verbos nem falsas-oxítonas. A primeira contém as palavras oxítonas em ordem alfabética, numeradas de 2 a 10.495. As colunas dois e três são as mesmas da lista-expandida, contendo o século arredondado (Data 1) e a data original do DH (Data 2), respectivamente. A quarta coluna grava a transcrição fonética de cada oxítone. A quinta traz o elemento final (abreviada como *E_F*) de cada palavra oxítone, ou seja, consoante, vogal ou glide, também transcritos foneticamente, os quais aparecem na sexta coluna como os números 0, 1 e 2, respectivamente significando consoante, vogal e glide. A sétima coluna apresenta a frequência-web (*F_Web*), tal como descrito acima. A oitava coluna mostra a língua que deu origem a palavras em português, no caso de empréstimos. Essa coluna também contém outras informações relevantes, como, por exemplo, “sem etimologia”, “(origem) controversa” etc. Por fim, a nona coluna apresenta a etimologia, tal como na lista-expandida. As colunas que marcavam a classe de palavras dos lexemas foram excluídas, posto que foram descartadas as palavras verbais e aquelas que fogem dos nossos objetivos.

Na lista-base, há cerca de 2.500 palavras sem descrição etimológica, o que dificulta sua classificação em relação à sua origem, embora, algumas vezes, a origem etimológica não seja de todo obscura ou desconhecida. É o caso, por exemplo, da palavra *nissei*, que está descrita no DH sem informação etimológica, ainda que saibamos que se trata de um empréstimo da língua japonesa. Mesmo sem essas informações, palavras que se enquadram nesse tipo foram incluídas na lista-base.

Adicionalmente, a lista-base contém palavras sem descrição etimológica e que, ao mesmo tempo, parecem ser ou são derivadas, ou

ainda são formadas a partir de falsos sufixos, como, por exemplo, é o caso da palavra *trilhão*, cuja etimologia pode ser associada, por analogia, às palavras *milhão* e *bilhão*.

DISCUSSÃO

Análise estatística dos dados

Nesta seção, analisaremos os dados referentes às colunas 2 e 8 da lista-base, que contêm o século arredondado e a língua de origem. Em seguida, analisaremos os dados referentes às colunas 5 e 6, que abarcam o elemento final das oxítonas, e, finalmente, analisaremos os dados referentes à coluna 7, que traz a frequência-web. É válido notar que, se a etimologia da palavra era controversa, escolheu-se a primeira língua citada pelo DH como possível emprestadora.

Além disso, na lista-base, foi adicionada uma categoria para topônimos, uma para antropônimos, uma para palavras derivadas de nomes próprios, uma para vocábulos expressivos, uma para formas históricas, uma para palavras resultantes de processo fonético-fonológico e uma última para palavras de difícil classificação, de acordo com a etimologia descrita no DH. Essas palavras serão levadas em conta apenas nas análises que não se referem à língua de origem, já que não possuem esta informação e, portanto, modificariam o resultado das análises estatísticas.

Análise da fonte emprestadora: língua e data

A Tabela 1, apresentada na página a seguir, contém os resultados da análise da etimologia das palavras oxítonas, destacando a origem etimológica, o número absoluto de entradas e também o valor relativo das quinze línguas³ que se revelaram mais representativas no *corpus*:

TABELA 1 - REPRESENTATIVIDADE NO *CORPUS*: LÍNGUAS EMPRESTADORAS

ORIGEM ETIMOLÓGICA	NÚMERO ABSOLUTO DE PALAVRAS OXÍTONAS NO DH	PORCENTAGEM EM RELAÇÃO AOTOTAL DE 10.494 PALAVRAS
latim	1774	16,9%
tupi	1558	14,9%
francês	583	5,6%
árabe iorubá	207	2,0%
português	190	2,0%
inglês	188	1,8%
origem indígena	157	1,5%
espanhol	140	1,3%
grego	87	0,8%
concani	78	0,7%
africanismo	74	0,7%
quimbundo	70	0,7%
italiano	65	0,6%
malaio	58	0,6%

Por um lado, o latim, o tupi,⁴ o francês, o árabe, o iorubá, o português, o inglês e o espanhol destacam-se com mais de 100 ocorrências. Por outro lado, um total de 76 línguas emprestou ao português quatro palavras ou menos, sendo que, dessas 76 línguas, 48 aparecem com somente um item lexical oxítono. O português foi desconsiderado como língua emprestadora, sendo sua presença na lista justificável por se tratar de palavras oxítonas cuja origem não pode ser associada aos processos morfológicos do português, embora também não haja evidências de se tratar de uma palavra estrangeira ou proveniente do latim, via evolução. Assim, as palavras de origem portuguesa foram desconsideradas. De um total de 10.494 palavras da lista-base, 81%, cerca de 8.500 palavras, são externas ao português e ao latim. No entanto, as palavras de origem

latina direta correspondem a 16,7% de todas as oxítonas. No que diz respeito às palavras de origem latina, é possível observar se as palavras latinas chegaram ao português por evolução ou por empréstimo tardio. No primeiro caso, a palavra passou por um processo gradativo de mudanças, através de fenômenos lingüísticos ocorridos desde o latim vulgar – idioma do qual o português é oriundo – até o português, como, por exemplo: *radix* > *raiz*, *aprīlis* > *abril* e *anellus* > *anel*. No segundo caso, a palavra passou por um processo erudito não-natural, que consistia em tomar palavras diretamente do latim e adaptá-las ao português, sobretudo no período Renascentista, como, por exemplo: *fūrōr* > *furor*, *beelzebūb* > *belzebu* e *trochaeus* > *troqueu*. Assim, se a palavra for datada como originária do século XIV em diante, tratar-se-á, possivelmente, de um empréstimo tardio. Dessa maneira, podemos analisar de modo mais detalhado as palavras de origem latina. Do total de palavras latinas (descritas acima), apenas 8%, ou seja, 145 palavras, não têm datação definida.

Cabe aqui uma palavra sobre o método de datação do DH. O método, não totalmente incontroverso, estabelece a datação das palavras de acordo com “a data do primeiro registro conhecido ou estimado de uma palavra, com indicação da fonte onde ocorre ou da primeira obra lexicográfica que a incluiu em sua nominata” (HOUAISS e VILLAR, 2001, p. XXI). Assim, as datas não são sempre precisas, pois a palavra já poderia estar sendo empregada na língua falada muito antes de ser efetivamente documentada. Ademais, 14,4% do total de entradas datadas da listabase são de palavras cuja datação é de 1899, que aparece 869 vezes (incluindo palavras do latim e de outras línguas). Neste caso, a fonte empregada no DH é a primeira edição (em dois volumes) do *Novo Dicionário da Língua Portuguesa*, de Cândido de Figueiredo. Portanto, reconhecemos que a datação dos dicionários etimológicos apresenta certos problemas.

Não somente o latim contribuiu com palavras oxítonas. As origens geográficas são as mais diversas, como podemos observar na Tabela 2,

que as divide em quatro grandes grupos: África, Américas, Europa (com três subgrupos: germânicas, latinas e outras) e Oriente (extremo e médio/próximo).⁵

TABELA 2 - GRANDES GRUPOS

GRUPO	NÚMERO ABSOLUTO	NÚMERO RELATIVO
África	522	9%
Américas	1783	29%
Europa 1 – Germânicas	215	4%
Europa 2 – Latinas	2797	46%
Europa 3 – Outras	109	2%
Oriente Médio e Próximo	552	9%
Extremo Oriente	78	1%

As línguas latinas da Europa aparecem com 46% do total das oxítonas incorporadas via empréstimos. As línguas das Américas, entre elas o tupi, aparecem com 29% do total. As línguas da África e as do Oriente (Próximo e Médio) aparecem em igual quantidade, 9% cada grupo. As oxítonas cujas origens podem ser remontadas às línguas do Extremo Oriente, às línguas germânicas e às outras línguas da Europa perfazem o total de 7% dos dados.⁶ Há, ainda, outras línguas de difícil classificação⁷ que foram descartadas, já que pouco relevantes, pois somam apenas 5 ocorrências, perfazendo menos de 1% dos dados.

Cada grande grupo pode ser analisado em separado. O grupo africano, por exemplo, representa 9% do total, comparecendo com 522 palavras; iorubá: 207 palavras, ou 40%; quimbundo: 70 palavras, ou 13 %; banto: 34 palavras, ou 7%; crioulo: 36 palavras, ou 7%; quicongo: 31 palavras, ou 6%; jeje: 18 palavras, ou 3%; umbundo: 10 palavras, ou 2%; outras⁸ línguas: 10 palavras; “africanismo”: 74 palavras, ou 14%; de origem africana: 24 palavras, ou 5%. Segundo Petter (2003), não há

um padrão para a classificação de africanismos nos dicionários de língua portuguesa, sendo possível observar insegurança na indicação de muitos vocábulos no DH. Nesse trabalho denominado “Palavras de origem africana nos dicionários *Houaiss* e *Novo Aurélio Século XXI*”, a autora procura analisar os termos definidos como africanismos no *Novo Aurélio Dicionário de Língua Portuguesa Século XXI* e no DH, buscando assim compreender a extensão desse conceito. Segundo Petter, “os termos de origem africana foram [...] constituindo uma classe importante entre os brasileirismos, distinguindo-se como africanismos” (PETTER, 2002, p. 107). Mas ainda há oscilações na definição e percepção desses termos. Petter cita os critérios propostos por Cunha (1987) para a definição de um brasileirismo, a saber: seu uso privativo no Brasil; seu uso exclusivamente regional; sua difusão junto a camadas sociais determinadas; a existência do termo em outra variante do português que não a brasileira e sua origem. Cita também as quatro espécies de brasileirismos propostas por Mattoso Câmara Junior (apud CUNHA, 1987, p. 27), uma das quais é o contato com outras línguas. No DH, aparecem as seguintes definições para *brasilismo*:

a) (Rubrica: lingüística) em sentido lato, qualquer fato de linguagem (fonético, morfológico, sintático, lexical, estilístico) próprio do português do Brasil; b) sob o ponto de vista lexical, palavra ou locução (dialetismo vocabular) ou aceção (dialetismo semântico) privativa do português do Brasil (HOUAISS e VILLAR, 2001, p. 507).

Já as definições de *africanismo* são:

a) (1858) (Rubrica: lingüística) palavra, construção ou expressão tomada de empréstimo de qualquer das línguas africanas; b) (Rubrica: lingüística) qualquer fato de linguagem (fonético, mórfico, sintático, lexical) privativo do português de alguma das ex-colônias portuguesas na África, em contraste com o de Portugal ou do Brasil; c) (Rubrica: lingüística) os fatos lexicais distintivos do português da África, não usados em Portugal ou no Brasil (HOUAISS e VILLAR, 2001, p. 107).

Segundo a definição acima, todos os empréstimos africanos seriam africanismos e seriam africanismos apenas expressões usadas no portu-

guês da África. Desse modo, as razões para o agrupamento entre “africanismo” e “de origem africana” são obscuras. Sobre isso, Petter (2003) afirma que, apesar da incoerência entre a definição da entrada *africanismo* e da consideração de *africanismo* como uma categoria identificadora da etimologia do termo, as palavras oriundas de uma língua africana em uso no PB não são classificadas como africanismos; são identificados como africanismos somente os termos próprios do português da África.

Outro problema de definição está no fato de que, segundo a definição de brasileirismo, africanismos não poderiam fazer parte da classe de brasileirismos como proposto por Petter e supracitado na definição do DH, posição diferente da adotada pelo *Novo Dicionário Aurélio de Língua Portuguesa Século XXI*.

Outra particularidade do DH, como observa Petter, é a utilização de um número mais amplo de fontes para determinar a etimologia dos vocábulos, usando, no entanto, apenas cinco fontes quando se trata de línguas africanas. Nele são usadas as seguintes denominações: “etimologia provavelmente africana”, “etimologia africana”, “de origem controversa”, “de origem obscura”, ou, então, menciona-se que um autor “sugere” um étimo, como no exemplo “orig.contrv.; Olga Cacciatore (apud HOUAISS e VILLAR, 2001) sugere o ior. *abi* ‘o que possui’ + *eko* ‘conhecimento’”. Segundo Petter, esse procedimento comprova que não houve nova pesquisa nessa área por parte dos lexicógrafos. A autora ainda diz que a análise desses dicionários “evidencia o conhecimento insuficiente do legado lexical das línguas africanas no português do Brasil” e, novamente, que “esse desconhecimento se explica em grande parte pela falta de pesquisa nova na área” (PETTER, 2003, p. 113).

Podemos ainda fazer um paralelo com outros termos utilizados nas definições de vocábulos no DH, como indigenismo e arabismo. Trataremos do último mais adiante. No caso de *indigenismo*, há a definição: a) (Rubrica: lingüística) palavra, construção ou locução de uma língua indígena tomada de empréstimo por outra língua (HOUAISS e

VILLAR, 2001, p. 1605), mas não a encontrada nos exemplos acima, que diz ser própria da língua a qual é oriunda. No segundo caso, temos essas duas definições: a) (Rubrica: lingüística) expressão característica da língua árabe; b) (Rubrica: lingüística) palavra, construção ou expressão próprias da língua árabe numa outra língua (HOUAISS e VILLAR, 2001, p. 169). Estes exemplos revelam que a insegurança na classificação dos vocábulos de origem africana no DH pode ser estendida para outras línguas, sobretudo à classificação inconsistente, mas amplamente empregada, de “indigenismo”.

Apesar desta imprecisão, de acordo com o DH, a língua africana que deu origem a um maior número de oxítonas foi o iorubá, com 207 ocorrências. Em 14% dos dados aparecem os africanismos, que correspondem a 13% dos dados. Essas palavras de origem supostamente iorubá se referem, em sua maioria, ao universo religioso e ritualístico,⁹ mas também a utensílios (ritualísticos ou não) e comida. Observemos alguns exemplos: *abô*: ior. {agbo} “infusão de mistura de folhas com fins medicinais”, segundo Cacciatore; *abroré*: segundo Cacciatore, ior. {ab}{-}{r}{è} “sacerdote chefe da floresta sagrada”; *adjá*: segundo Cacciatore, do ior. {aadja} “tipo de chocalho us. em cerimônias rituais”, der. de {dja} “bater”; *oim*: ior. {oyin} “abelha; mel”; *orô*: ior. {oro} “costume, tradição”.

No que diz respeito à origem americana das oxítonas, a língua tupi aparece como a maior influenciadora comparecendo com 1.558 ocorrências. 16 palavras (1%) são descritas como de origem guarani. Palavras de origem indígena indeterminada aparecem em 12% dos dados, sendo a somatória de “origem indígena” de 157 palavras (9%), “indigenismo”: 24 palavras (1%), e “outras línguas indígenas”: 27 palavras (2%). Este dado é relevante, pois mostra que mais de 10% das palavras atribuídas ao grupo americano são de origem indígena, embora sua filiação lingüística seja desconhecida e que todas as outras línguas indígenas americanas contribuíram com somente 2% do total do grupo. Graças ao avanço nos estudos das línguas indígenas brasileiras nos últimos trinta anos, esse problema poderia estar mais perto de ser

solucionado, pois se coloca como uma questão central para os especialistas nas línguas indígenas brasileiras: qual a extensão da influência vocabular indígena no português, excluindo-se as palavras de origem tupi? A maior parte das palavras do tupi se refere à nomenclatura de fauna e flora. Observemos alguns exemplos: *abati*: tupi {awa'ti} “milho, arroz, trigo”; *acari*: tupi {gwaka'ri} “peixe da fam. dos loricariídeos, tb. conhecido como cascudo”; *cacundê*: tupi {ka'a} “folha, mato, erva” + {ku'ndawa} “retorcido”; *imburi*: tupi {mbu'ri} “espécie de palmeira”.

O latim e as línguas neolatinas formam um grupo influente.¹⁰ O latim aparece em 1.774 ocorrências (63%) dos dados do grupo das línguas latinas. Com 21% temos o francês, com 583 palavras, o espanhol com 140 palavras e o italiano com 65.

O grupo de línguas germânicas contribuiu com 215 palavras, sendo que o inglês aparece com 188 palavras (88%) do total do grupo. O alemão (20 palavras), o dinamarquês (1 palavra), o germânico (3) e o gótico (3) são as outras línguas do grupo.

Outras línguas da Europa tiveram uma influência menos marcante, excetuando-se o grego com 87 palavras. As outras línguas do grupo são o basco (2), o russo (7), o turco (10) e o húngaro (1). As línguas do grupo do Extremo Oriente que incluem o birmanês, o chinês, o japonês, o mongol e o vietnamita aparecem com menos de 80 palavras, com destaque para o chinês com 37 e o japonês com 36 palavras.

No grupo das línguas de origem do Oriente Próximo e Médio, o árabe destaca-se com 207 palavras, 38% do total do grupo. É válido ressaltar que os arabismos, as palavras de origem moçárabe e as palavras do persa-árabe não foram computadas nesses 38%, mas sim nos 12% de *outras*. Em relação aos arabismos, é difícil determinar se a palavra é um empréstimo do árabe ou uma palavra portuguesa influenciada pelo árabe, ou seja, arabizada. Já as palavras de origem moçárabe, segundo o próprio DH, são originárias de dialetos românicos falados pelos moçárabes e não da língua árabe. As palavras do persa-árabe também não devem ser consideradas como da língua árabe porque é difícil comprovar sua

origem, já que essa classificação engloba duas línguas tipologicamente distintas. Esse grupo ainda contém o persa (com 35 palavras, 6% do total do grupo), o sânscrito (28), o hebraico (27), o concani (78), o concani-marata (16), o hindi (22), o malaio (58) e um grupo formado por outras línguas minoritárias¹¹ com 69 palavras. A forte presença concani e concani-marata (embora tipologicamente distintas, apresentadas no DH como um grupo homogêneo) deve-se à influência do concani no vocabulário português de Goa, Índia.

A quantidade de palavras oxítonas originárias das nove línguas majoritárias (latim, tupi, francês, árabe,¹² iorubá, inglês, espanhol, grego e concani) será correlacionada com a data da entrada do étimo no português. Primeiramente, é preciso separar o número de palavras com e sem datação. Como isto já foi feito acima com o latim, apresentaremos a Tabela 3, contendo a porcentagem de empréstimos com datação definida e indefinida e seus respectivos valores absolutos e relativos associados às línguas restantes.

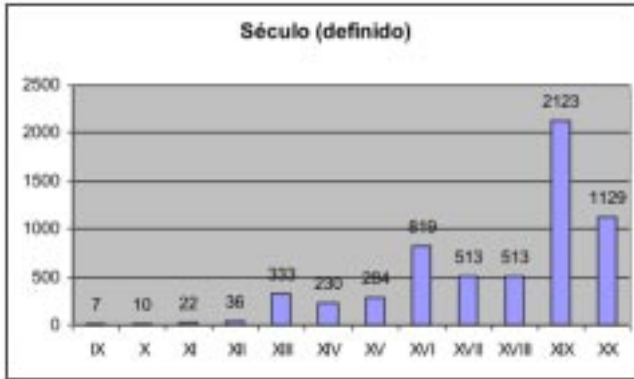
TABELA 3 - DATAÇÃO DEFINIDA E INDEFINIDA

	COM DATAÇÃO (VALOR ABSOLUTO)	COM DATAÇÃO (VALOR RELATIVO)	SEM DATAÇÃO (VALOR ABSOLUTO)	SEM DATAÇÃO (VALOR RELATIVO)
Árabe	180	87%	27	13%
Tupi	928	60%	630	40%
Francês	459	79%	124	21%
Grego	72	83%	15	17%
Concani	68	87%	10	13%
Espanhol	120	86%	20	14%
Iorubá	135	65%	72	35%
Inglês	117	62%	71	38%

Na lista-base, 57% das palavras oxítonas têm datação precisa. Das palavras com datação precisa, 35,3% entraram no século XIX. Em todos os outros séculos a entrada foi bem menor, mas vale destacar também que

no século XX houve um influxo de cerca da metade do número de palavras que entraram no século XIX, como observamos na Figura 1.

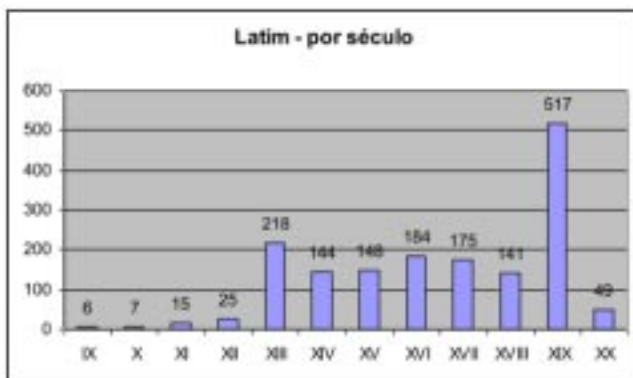
FIGURA 1 - OXÍTONAS POR SÉCULO



A seguir, apresentaremos dados relativos ao número absoluto de empréstimos de palavras oxítonas por século de cada uma das línguas restantes¹³ (as de maior peso nos empréstimos) e explicitaremos também os valores relativos. A quantidade de palavras oxítonas de origem latina que entrou é, segundo o DH, a seguinte: 6 palavras oxítonas no século IX e 7 no século X, atingindo um número estatisticamente insignificante; 15 palavras (1%) no século XI; 25 palavras (2%) no século XII; 218 palavras (13%) no século XIII; 144, 148 e 141 palavras ou aproximadamente 11% (em cada século) nos séculos XIV, XV e XVIII, respectivamente; 517 palavras (32%) no século XIX e 49 palavras (3%) no século XX. Exemplifiquemos algumas entradas para cada século a título de ilustração. Das entradas por evolução, temos: no século IX: *casâlis* > *casal*; no século X: *falcōnis* > *falcão*; no século XI: *judeu* > *judaeos*; no século XII: *retro* > **redro* > *redor*; e, finalmente, no século XIII: *albōris* > *alvor*. Via empréstimo tardio, temos: no século XIV: *avellāna* > *avelã*; no século XV: *capax* > *capaz*; no século XVI: *anthrax* > *antraz*; no século XVII: *athēus* > *ateu*; no século XVIII: *archōgos* > *arqueu*; no século XIX: *turōnis* > *turão*; e no século XX: *turgōris* > *turgor*. Os dados sugerem um pico inegável no século XIX. Observamos que a

entrada tardia é constante, pois a maior parte das palavras oxítonas oriundas do latim é empréstimo no português, ou seja, não entrou no português por evolução.

FIGURA 2 - OXÍTONAS ORIUNDAS DO LATIM



Para o “tupi”, os valores são: 197 palavras (24%) no século XVI; 84 e 80 palavras (10%) nos séculos XVII e XVIII, cada; 301 palavras (37%) no século XIX e 19% no século XX. O pico inicial na primeira fase da colonização era esperado. No entanto, houve uma entrada substancial no século XIX que pode estar associada à documentação e não mais à influência direta das línguas do grupo tupi.

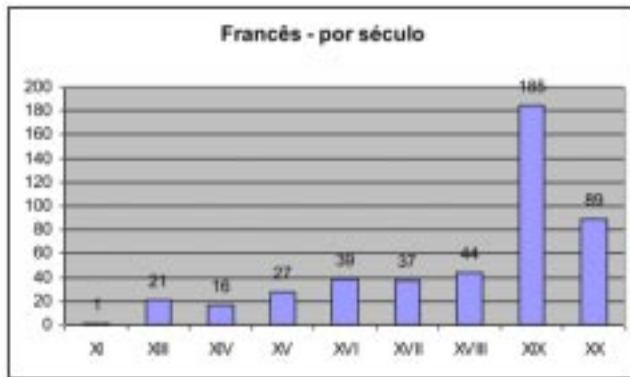
FIGURA 3 -OXÍTONAS ORIUNDAS DO “TUPI”



Já o francês apresenta seu pico nos séculos XIX e XX, graças, sobretudo, à influência cultural francesa.

Os valores são os seguintes: 1 palavra no século XI; não há registros no século XII; 21 palavras (5%) no século XIII; 16 palavras (3%) no século XIV; 27 palavras (6%) no século XV; 39 palavras (8%) nos séculos XVI e XVII, cada; 44 palavras (10%) no século XVIII; 185 palavras (41%) no século XIX e 89 palavras (19%) no século XX. Ou seja, houve um influxo regular, porém limitado, nos séculos anteriores aos séculos XIX e XX.

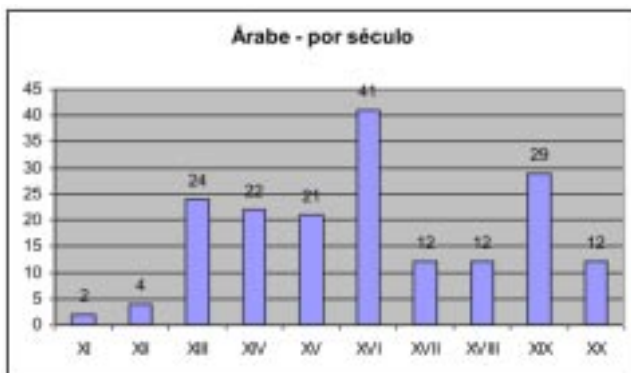
FIGURA 4 - OXÍTONAS ORIUNDAS DO FRANCÊS



Quanto ao árabe, a documentação analisada mostra que a influência diminuiu depois do século XVI, embora no século XIX tenha havido um pico, juntamente com o francês, o grego, o espanhol, o latim e o tupi com forte presença no século XIX.

Os valores relativos referentes ao árabe são: 2 palavras (1%) no século XI; 4 palavras (2%) no século XII; 24 palavras (13%) no século XIII; 22 palavras (12%) nos séculos XIV e XV, cada; 41 palavras (23%) no século XVI; 12 palavras (7%) nos séculos XVII e XVIII, cada; 29 palavras (16%) no século XIX e 12 palavras (7%) no século XX.

FIGURA 5 - OXÍTONAS ORIUNDAS DO ÁRABE



Os empréstimos oriundos do iorubá apresentam o seguinte quadro: 2 palavras (1%) no século XVI; 27 palavras (20%) no século XIX e 106 palavras (79%) no século XX. A lacuna nos séculos XVII e XVIII pode ser atribuída à escassez de documentação e ao desinteresse pela cultura iorubá e suas religiões.

FIGURA 6 - OXÍTONAS ORIUNDAS DO IORUBA



Os valores relativos referentes às palavras de origem inglesa: 3 palavras ou cerca de 3% nos séculos XVI e XVIII, cada; 22 palavras (19%) no século XIX e 88 palavras (75%) no século XX, mostram também destacada influência no século XX, associado ao prestígio cultural e econômico dos países de língua inglesa nesses dois últimos séculos.

FIGURA 7 - OXÍTONAS ORIUNDAS DO INGLÊS



Já o espanhol apresenta uma influência que perpassa os séculos. Sobretudo no período que vai do século XIV ao XIX: 2 palavras nos dois primeiros séculos, uma em cada; 6 palavras (5%) no século XIII; 10 palavras (8%) no XIV e a mesma quantidade no XV; 26 palavras (22%) no século XVI; 23 palavras (19%) no século XVII; 14 palavras (12%) no século XVIII; 27 palavras (22%) no século XIX e somente duas palavras (2%) no século XX.

FIGURA 8 - OXÍTONAS ORIUNDAS DO ESPANHOL



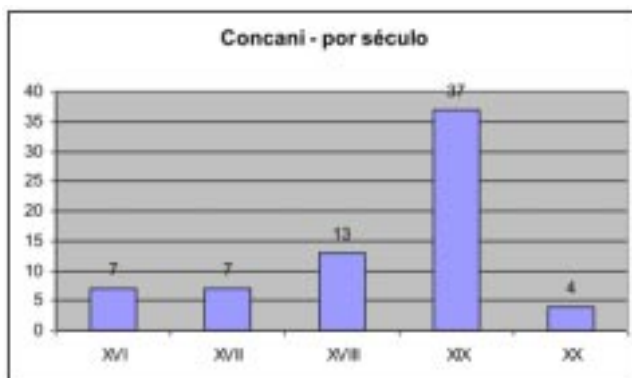
Diferentemente do latim, o grego apresenta uma influência limitada, pelo menos no que diz respeito às palavras oxítonas, embora seu pico seja no século XIX: 3 palavras (4%) no século XIII; uma palavra (1%) no século XIV; 4 palavras (6%) no século XV; nos séculos XVI e XVIII 9 palavras (13%) em cada século; 28 palavras (38%) no século XIX e 7 palavras (10%) do total no século XX.

FIGURA 9 - OXÍTONAS ORIUNDAS DO GREGO



A presença do concani no grupo das línguas que mais deram empréstimos ao português não é totalmente estranha se considerarmos que o concani era a língua falada na região de Goa, na Índia. Goa deixou de ser colônia portuguesa somente em 1961 e a presença portuguesa foi mais marcante em Goa do que em outras possessões do oriente. Além disso, a administração portuguesa influenciou a região de Goa e foi influenciada pela cultura concani. Outras palavras oxítonas de origem supostamente concani como as das línguas concani-marata, marata-concani (DH) e concani-guzarate não estão computadas na Figura 10. Os valores relativos referentes ao concani (Figura 10) são: 7 palavras (10%) nos séculos XIV e XVII, cada; 13 palavras (19%) no século XVIII; 37 palavras (55%) no século XIX e 4 palavras (6%) no século XX.

FIGURA 10 - OXÍTONAS ORIUNDAS DO CONCANI



A partir dos dados apresentados nesta seção, estabelecemos o século de maior incidência de empréstimos de cada uma dessas línguas e a porcentagem de empréstimos no dado século, relativamente ao total existente.

TABELA 4 - SÉCULO DE MAIOR INCIDÊNCIA

	SÉCULO DE MAIOR INCIDÊNCIA DE EMPRÉSTIMOS	PORCENTAGEM DE EMPRÉSTIMOS NESSE SÉCULO EM RELAÇÃO AO TOTAL ATUALMENTE ENCONTRADO
Espanhol	XVI e XIX	22% em cada século
Árabe	XVI	22%
Concani	XIX	55%
Francês	XIX	41%
Grego	XIX	38%
Tupi	XIX	37%
Latim	XIX	32%
Iorubá	XX	79%
Inglês	XX	75%

Assim, podemos formar o seguinte quadro-resumo:

Do total de 87% das palavras do árabe com datação, 22% entraram no século XVI;

Do total de 92% das palavras do latim com datação, 32% entraram no século XIX;

Do total de 60% das palavras do tupi com datação, 37% entraram no século XIX;

Do total de 79% das palavras do francês com datação, 41% entraram no século XIX;

Do total de 83% das palavras do grego com datação, 38% entraram no século XIX;

Do total de 87% das palavras do concani com datação, 55% entraram no século XIX;

Do total de 86% das palavras do espanhol com datação, 22% entraram no século XVI e 22% no XIX;

Do total de 65% das palavras do iorubá com datação, 79% entraram no século XX;

Do total de 62% das palavras do inglês com datação, 75% entraram no século XX.

Análise do elemento final das oxítonas

Nesta seção, abordaremos os elementos finais das oxítonas, através dos dados das colunas 5 e 6 da lista-base, que são o elemento final e sua classificação.

Primeiramente, separaremos os dados do elemento final da palavra, considerando se são consoante, vogal, glide ou vogal nasalizada, a fim de determinar sua porcentagem, para mais adiante examinar esses elementos individualmente. Verificamos a seguinte distribuição dos elementos finais nas oxítonas: as vogais são o elemento final de 50% das oxítonas, ou seja, estão presentes em 5.273 palavras. As vogais nasalizadas aparecem em 1.225 palavras, ou 12% dos dados. Já os glides [w] e [j] aparecem em 26%, ou 2.685 palavras, enquanto que as consoantes aparecem em 1.311 das palavras, ou 12%. As vogais que ocorrem com

maior frequência como elemento final nas oxítonas são [i], 25%, ou 1.273 palavras; [a], 24%, 1.255 palavras; e [u], 22% ou 1.162 palavras, enquanto que a vogal [ɛ] comparece com 11% ou 596 palavras, a vogal [ɔ] comparece com 8% ou 444 palavras, e a vogal [e] aparece em 7% ou 368 palavras, sendo as menos frequentes. A vogal menos frequente é o [o], com 3%, ou apenas 175 ocorrências.

No que diz respeito aos glides, o elemento [w] aparece em 97% das oxítonas terminadas em glides, ou seja, em 2.597 ocorrências, enquanto que o [j] aparece em apenas 3% dos dados, ou seja, em 86 ocorrências.

Segundo Silva (2006), a nasalidade é regressiva no português, ou seja, as vogais e ditongos nasais do português resultam de vogais seguidas de consoantes nasais no latim. Temos, por exemplo: *româna* > *romã*; **burricana* > *borrega*; *avellâna* > *avelã*; **abigone* > *abegom* > *abigão*; *matiâna* > *maçaam* > *maçã*; *hortulâna* > *hortelaam* > *hortelã*; *jejûnus* > *jejum*; *cherubîm* > *querubim*. Podemos observar, nos exemplos acima, a nasal latina /n/ em posição intervocálica, vogais e ditongos nasais em posição final de vocábulo e vogais seguidas de nasal implosiva. Dessa forma, as vogais nasalizadas [ĩ] e [ã] são as que aparecem na maior parte dos dados, com 38% (468 palavras) e 31% (385 palavras), respectivamente. A vogal [ũ] aparece com 13%, 154 ocorrências, a vogal [ẽ] com 10%, ou 117 palavras e, finalmente, [õ] com 8%, ou 101 palavras.

Consideramos duas consoantes [r] e [s] e descartamos as consoantes ortográficas [m], [e] e [z], pois, em todos os casos, não ocorrem como elementos da coda quando a palavra é pronunciada isoladamente. No primeiro caso, ocorre a nasalização da vogal e o desaparecimento da consoante *m*; no segundo, a consoante em coda *l* é pronunciada como o glide [w] (na maioria dos dialetos do Brasil, incluindo o paulista), e, no terceiro, a consoante *z* é pronunciada como [s]. Avaliando somente *r* e *s*, verificamos que o [r] aparece em 71% dos dados, ou seja, em 928 palavras, sendo mais frequente do que o [s], que aparece em 29% dos casos, ou em 385 palavras.

O quadro a seguir apresenta [s] a distribuição das palavras com base no elemento final:

w	2.597	24,71%
i	1.273	12,13
a	1.255	11,96
u	1.162	11,07
r	928	8,85
ε	596	5,68
ĩ	468	4,46
o	444	4,24
ã	385	3,68
s	385	3,68
e	368	3,50
o	175	1,68
ũ	154	1,47
ẽ	117	1,11
õ	101	0,96
j	86	0,82
Total	10.494	100%

Avaliando todas as oxítonas, o elemento final mais freqüente é glide [w], com 24,71% to total, ou seja, de cada quatro oxítonas, uma termina com este glide. Do outro lado da tabela está o glide [j], com apenas 86 ocorrências ou 0,82% do total. Cada uma das vogais /i/, /a/ e /u/ representa cerca de 11% do total das oxítonas.

Análise da freqüência

A freqüência-web indica cada ocorrência individual de uma palavra oxítona em cada página da internet através do site de buscas Google. Dessa forma, é um indicador da freqüência geral na internet, o que, por hipótese, demonstra quão comum a palavra é. Nessa consulta 34% das oxítonas apareceram em mais de 1000 páginas.

Utilizando os critérios estabelecidos por Araújo et. al. (2007, p.23) para a delimitação de freqüência de palavras, denominamos *raras* as

palavras que ocorriam em menos de 10 páginas, *incomuns* aquelas encontradas em até 200 páginas, *comuns* as que ocorriam em até 20.000 páginas e *freqüentes* as que foram encontradas em mais de 20.000 páginas.

Observamos que 48,1% das oxítonas são *raras* ou *incomuns* e 51,9% são *comuns* ou *freqüentes*. Ao sobrepor estes resultados aos de Araújo et. al. (2007, p.23), verificamos uma ínfima diferença nos resultados. Segundo esses autores, 48,3% das oxítonas são *raras* ou *incomuns* e 53,8% são *comuns* ou *freqüentes*, número idêntico ao das oxítonas da lista-expandida. As conclusões de nossa pesquisa ficaram muito próximas do resultado obtido por Araújo et. al (2007) que, através do exame de um *corpus* geral de 150.875 palavras do DH (0,4% monossílabos, 24,9% oxítonas,¹⁴ 62,5% paroxítonas e 12,2% proparoxítonas), concluiu que 53,8% são *raras* ou *incomuns*, enquanto que 46,3% são *comuns* ou *freqüentes*. Segundo Cintra (1997), as paroxítonas, que se constituem no padrão de acentuação mais freqüente do português, representam 42% dos dados analisados pelo autor e 62,5% dos dados do *corpus* geral de Araújo et al. Assim, podemos comparar a freqüência relativa (em relação ao seu total) das paroxítonas com a das oxítonas. Observamos esses dados com mais detalhes na Tabela 5:

TABELA 5 - FREQUÊNCIA-WEB

	RARAS	INCOMUNS	COMUNS	FREQUENTES
Lista-base (10.494 oxítonas)	23,8%	24,3%	26,8%	25,1%
Lista-expandida (37.591 oxítonas)	21,1%	27,2%	26,1%	25,6%
Oxítonas do corpus de Araújo et al. (37.568 oxítonas)	21,1%	27,2%	26,1%	25,6%
Paroxítonas do corpus de Araújo et al. (94.297 paroxítonas)	26,3%	26,4%	24,2%	23,1%
Corpus geral de Araújo et al. (150.875 palavras)	27,5%	26,3%	23,7%	22,6%

Os dados mostram que não foram excluídos os verbos e as palavras formadas por derivação e composição do *corpus* de oxítonas utilizado por Araújo et. al (2007). Apesar disso, os resultados não foram comprometidos. Observamos também que oxítonas e paroxítonas têm aproximadamente a mesma frequência relativa, ou seja, a mesma frequência em relação ao seu respectivo número total. Das 10.494 oxítonas da lista-base, 51,9% são comuns ou freqüentes, assim como 47,3% do *corpus* de 94.297 paroxítonas de Araújo et. al. O mesmo ocorre com o *corpus* geral de Araújo et. al. (2007), das 150.875 palavras, 46,3% são comuns ou freqüentes.

CONSIDERAÇÕES FINAIS

O objetivo do texto era mostrar que, baseando-se em critério etimológico, a maioria das palavras nominais oxítonas podem ser classificadas como empréstimo. De certa forma, a existência de palavras com acento final no português é uma inovação em relação ao latim, que não o permitia. Assim, as palavras com acento final em português podem ser derivadas das evoluções fonéticas do latim, dos processos internos ao português, ou da adaptação de empréstimos de várias línguas. Estas adaptações revelam que o padrão oxítono é uma característica da língua portuguesa desde os seus primórdios e não há nenhuma restrição que proíba a entrada de novas palavras oxítonas no português. Portanto, esse padrão ainda é produtivo.

A análise estatística dos elementos finais mostra também que cerca de 50% (ou 5.254 palavras) das oxítonas terminam em vogais. Isso é um desafio para as teorias do acento em português (ARAÚJO, 2007), que defendem que o peso silábico é importante para se estabelecer o *locus* do acento. Outra conclusão deste estudo diz respeito à frequência das oxítonas: a análise dos dados sugere que as oxítonas são, em média, tão comuns quanto as palavras paroxítonas e proparoxítonas.

O estudo também sugere que há uma relação entre o grau de influência cultural ou econômica de uma determinada língua e o número de empréstimos. Esta hipótese, contudo, precisa ser comprovada, levando-se em conta os empréstimos de palavras paroxítonas e proparoxítonas.

No entanto, uma outra questão permanece em aberto e requer novas pesquisas: a relação entre a posição de acento original na língua emprestadora e a efetiva adaptação em português e suas conseqüências. Este estudo requererá, entre outras coisas, um conhecimento dos sistemas prosódicos das principais línguas emprestadoras e as possibilidades de adaptação ao longo da história do português. Como ponto de partida, seria interessante comparar as evoluções fonéticas ocorridas na mudança do latim para o português com o contato entre o português e línguas com acento final (como o francês e a maioria das línguas do grupo tupi) e até mesmo línguas cujo sistema acentual seja diverso do português ou das línguas neolatinas, como é o caso do sistema *pitch-accent* do japonês, por exemplo.

A STATISTICAL ACCOUNT OF FINAL STRESS IN PORTUGUESE

ABSTRACT

The aim of this paper is to analyse a *corpus* of final stress words (also known as oxytone) in Portuguese. Based on a *corpus* of 10.494 noun words with final stress and their respective phonetic transcriptions from the Houaiss Dictionary, we argue that the majority of the final stress words in Portuguese are lexical borrowings. Latin, “Tupi”, French, Arabic, Ioruba are the majors sources, however, Portuguese borrowed from more than one hundred languages. We also discuss the quality of the final element (if vowel, glide or consonant) and the overall frequency of these final stress words.

KEY WORDS: final stress, portuguese, phonology, oxytone.

NOTAS

- 1 Para isso recorremos ao pós-doutorando Dr. Zwinglio Guimarães-Filho e sua equipe de colaboradores do Instituto de Física da USP, a partir da sugestão

do Prof. Dr. Mário Eduardo Viaro da FFLCH. A equipe do Dr. Guimarães-Filho nos auxiliou na formatação da base de dados que constitui o corpus da pesquisa.

- 2 A lista-expandida foi organizada em colunas. A primeira coluna contém as palavras oxítonas numeradas de 1 a 37.591, em ordem alfabética, conforme o DH. A segunda coluna contém, a partir da datação do dicionário, o século arredondado, ou seja, para fins estatísticos, o ano preciso da datação foi normalizado para uma data cheia. No caso de datação imprecisa ou desconhecida, atribuiu-se o valor zero. Com esta informação, pode-se também determinar se uma palavra latina possui seu primeiro registro de datação no português por evolução ou por empréstimo tardio. Neste caso, por exemplo, se a palavra for datada como oriunda do século 14 em diante, será um empréstimo erudito. Posteriormente, isto nos permitirá também estabelecer em quais séculos houve maior entrada das oxítonas no português e se houve uma maior incidência de determinada língua específica em algum século. A terceira coluna contém a data original registrada no dicionário e sua respectiva fonte. A datação precisa pode ser útil para estudos sócio-culturais, estabelecendo, por exemplo, a maior influência de uma determinada língua estrangeira em um período qualquer. A quarta coluna contém a transliteração, gerada a partir dos dados do programa do DH e da programação, para tal fim, no MatLab. Na lista final, esta coluna foi substituída pela transcrição fonética, feita manualmente, considerando-se o dialeto paulistano. A quarta coluna apresenta a sílaba final da palavra, sendo importante para verificarmos a sua constituição, sobretudo sua vogal e consoante finais, as quais aparecem na quinta e na sexta colunas, respectivamente. A sétima coluna é um desdobramento das colunas 5 e 6. Foram atribuídos os números zero e 1, respectivamente para vogal e consoante, com o intuito de composição e análise dos dados. A oitava coluna apresenta a frequência-web, que consiste na análise da frequência de uso das palavras oxítonas, considerando-se sua ocorrência em páginas em português da internet indexadas pelo site de buscas Google, www.google.com.br. A frequência indica cada ocorrência individual de uma palavra oxítona em cada página. Dessa forma, a frequência-web é um indicador da frequência geral na internet. As colunas 9 e 10 marcam a classe de palavra do lexema, separando-as entre verbo e não-verbo (substantivos, adjetivos, advérbios etc.). Para tanto, a legenda empregada nestas duas colunas foi 1 para sim e zero para não. A décima primeira coluna foi inserida depois que a lista-expandida

estava pronta. Cada uma das palavras foi analisada, a fim de classificá-la apta à lista-base, a partir dos critérios supramencionados. Por fim, há a coluna com a informação etimológica, a qual traz informações históricas, morfológicas e também sobre os processos fonológicos do processo de nativização.

- 3 Demais línguas: chinês, 37 palavras, 0,4% do total das oxítonas; crioulo, 36, 0,3%; japonês, 36, 0,3%; persa, 35, 0,3%; (de origem) banta, 34, 0,3%; quicongo, 31, 0,3%; sânscrito, 28, 0,3%; hebraico, 27, 0,3%; (de origem) africana, 24, 0,2%; indigenismo, 24, 0,2%; hindi, 22, 0,2%; alemão, 20, 0,2%; jeje, 18, 0,2%; concani-marata, 16, 0,2%; guarani, 16, 0,2%; provençal, 16, 0,2%; marata, 13, 0,1%; guzarate, 11, 0,1%; turco, 10, 0,1%; umbundo, 10, 0,1%; catalão, 8, 0,1%; hispano-americano, 8, 0,1%; neo-árabe, 8, 0,1%; russo, 7, 0,1%; (de origem) caribe, 6, 0,1%; tamil, 6, 0,1%; cariri, 5, 0,1%; asteca/náuatle, 4, 0,0%; cingalês, 4, 0,0%; hauçá, 4, 0,0%; hispano-árabe, 4, 0,0%; nheengatu, 4, 0,0%; bangali, 3, 0,0%; galibi, 3, 0,0%; germânico, 3, 0,0%; gótico, 3, 0,0%; malaio-javanês, 3, 0,0%; marata-concani, 3, 0,0%; qechua, 3, 0,0%; tamil-malaio, 3, 0,0%; (de origem) ameríndia, 2, 0,0%; arabismo africano, 2, 0,0%; basco, 2, 0,0%; birmanês, 2, 0,0%; dravídico, 2, 0,0%; espanhol platino, 2, 0,0%; franco, 2, 0,0%; javanês, 2, 0,0%; kwa, 2, 0,0%; macua, 2, 0,0%; marata-guzarete, 2, 0,0%; mongol, 2, 0,0%; persa-árabe, 2, 0,0%; tapuio, 2, 0,0%; tupi-guarani, 2, 0,0%; (de origem) altaica, 1, 0,0%; amárico, 1, 0,0%; arabismo, 1, 0,0%; armênio, 1, 0,0%; aruaque, 1, 0,0%; (de origem) bengali, 1, 0,0%; berbere, 1, 0,0%; butanês, 1, 0,0%; (de origem) céltica, 1, 0,0%; (de origem) cigana, 1, 0,0%; cigano, 1, 0,0%; concani-guzarati, 1, 0,0%; copta, 1, 0,0%; crioulo macaense, 1, 0,0%; dinamarquês, 1, 0,0%; divehi, 1, 0,0%; egípcio, 1, 0,0%; (de origem) escandinava, 1, 0,0%; fula, 1, 0,0%; galg, 1, 0,0%; ganguela, 1, 0,0%; havaiano, 1, 0,0%; hindi-persa, 1, 0,0%; húngaro, 1, 0,0%; indo-iraniano, 1, 0,0%; jê, 1, 0,0%; língua de Moçambique, 1, 0,0%; lun, 1, 0,0%; lutiazi, 1, 0,0%; malgaxe, 1, 0,0%; malinquê, 1, 0,0%; marata-hindustâni, 1, 0,0%; marati, 1, 0,0%; (de origem) moçárabe, 1, 0,0%; nupê, 1, 0,0%; occitano, 1, 0,0%; panjabi, 1, 0,0%; pareci, 1, 0,0%; purocoto, 1, 0,0%; semítica, 1, 0,0%; tagalog, 1, 0,0%; tsonga, 1, 0,0%; twi, 1, 0,0%; uigur, 1, 0,0%; vietnamita, 1, 0,0%; e xipaia, 1, 0,0%. As línguas com porcentagem de 0,0% são estatisticamente insignificantes.
- 4 O uso do termo “tupi” segue tão somente a nomenclatura do DH. A língua tupi, efetivamente, nunca existiu. A expressão “língua tupi” refere-se ao

dialeto empregado pelos jesuítas, sem especificação das diferenças regionais, muito mais do que a uma língua específica. Cientificamente, emprega-se o termo “tupi” como denominação de um tronco lingüístico, embora, nos círculos não-científicos e paracientíficos, empregue-se o termo “tupi” como designação de uma língua indígena brasileira (RODRIGUES, 1986).

- 5 *Oriente Próximo/Oriente Médio*: amárico, árabe, arabismo, arabismo africano, armênio, bengali, butanês, concani, cingalês, concani-guzarati, concani-marata, copta, crioulo macaense, divehi, dravídico, guzarete, hebraico, hindi, hindi-persa, indo-ariano, javanês, malaio, malaio-javanês, marata, marata-concani, marata-guzarete, marata-hindustâni, marati, neo-árico, origem bengala, origem moçárabe, panjabi, persa, persa-árabe, sânscrito, semítica, tagalog, tamil, tamil-malaio e uigur. *África*: africanismo, berbere, crioulo, egípcio, fula, ganguela, hauçá, iorubá, jeje, kwa, língua de Moçambique, lun, macua, malgaxe, nupê, origem africana, origem banta, quicongo, quimbundo, tsonga, umbundo e twi; *Américas*: asteca/náuatle, cariri, galabi, guarani, havaiano, indigenismo, jê, nheengatu, origem ameríndia, origem caribenha, origem indígena, parecí, quechua, tapuio, tupi, tupi-guarani e xipaia; *Europa* (germânicas): alemão, dinamarquês, germânico, gótico e inglês; *Europa* (latinas): catalão, espanhol, espanhol platino, francês, franco, hispano-americano, hispano-árabe, italiano, latim, occitano, português e provençal; *Europa* (outras): basco, grego, húngaro, origem céltica, origem escandinava, russo e turco; *Extremo Oriente*: birmanês, chinês, japonês, mongol e vietnamita;
- 6 É válido ressaltar que toda a classificação das línguas nos grupos acima citados foi feita através do próprio DH.
- 7 A saber: galg, luziati, malinquê, origem cigana e purocoto.
- 8 Foram colocadas como “outras”, as línguas com menos de 10 ocorrências, somando 3% dos dados, sendo elas: berbere, egípcio, fula, ganguela, hauçá, kwa, língua de Moçambique, lun, macua, malgaxe, nupê, tsonga, e twi.
- 9 Petter (2003) afirma que “como consequência dessa expansão de fontes consultadas vai haver um maior número de entradas de origem africana, relativas, principalmente, ao universo religioso, tendo como fonte Cacciatore”.
- 10 As outras línguas deste grupo são: catalão, espanhol platino, franco, hispano-americano, hispano-árabe e occitano. Com menos de 8 ocorrências, somam 1% dos dados.

- 11 Amárico, arabismo, arabismo africano, armênio, bangali, butanês, cingalês, concani-guzarati, copta, crioulo macaense, divehi, dravídico, guzarete, hindipersa, indo-ariano, javanês, malaio-javanês, marata-concani, marata-guzarete, marata-hindustâni, marati, neo-árabe, origem bengala, origem moçárabe, panjabi, persa-árabe, semítica, tagalog, tamil, tamil-malaio e uigur. Todas com menos de 12 ocorrências, somando 12% dos dados.
- 12 Os números relativos ao árabe excluem os arabismos, as palavras de origem moçárabe e persa-árabe.
- 13 A saber: tupi, francês, árabe, iorubá, inglês, espanhol, grego e concani.
- 14 Ou seja, 37.568 oxítonas (número próximo ao da lista-expandida: 37.591).

REFERÊNCIAS

- AGOSTINHO, A. L. S. *Relatório do projeto palavras oxítonas em português*. São Paulo: FFLCH-USP, 2007. (Mimeografado).
- ARAÚJO, G. (Org.). *O acento em português: abordagens fonológicas*. São Paulo: Parábola, 2007.
- ARAÚJO, G. A. et al. Algumas observações sobre as proparoxítonas e o sistema acentual do português. In: ARAÚJO, G. (Org.). *O acento em português: abordagens fonológicas*. São Paulo: Parábola, 2007.
- CACCIATORE, Olga G. *Dicionário de cultos afro-brasileiros*. Rio de Janeiro: Forense Universitária, 1988.
- CAVAZZI, G. A. *Descrição histórica dos três reinos do Congo, Matamba e Angola, de João Antonio Cavazzi de Montecúcolo*. Lisboa: Junta de Investigação do Ultramar/ Agrupamento de estudos de cartografia antiga, [1687]1965.
- CINTRA, G. Distribuição de padrões acentuais no vocábulo em português. *Confluência*, n. 5/3, Unesp/Assis, p. 82-93, 1997.
- CORREA, E. A. da S. *Historia de Angola*. Lisboa: Atica, [1782]1937, 2v.
- CUNHA, C. O. *O que é um brasileiro?* Rio de Janeiro: Tempo Brasileiro, 1987.
- FERREIRA NETTO, W. *Introdução à fonologia da língua portuguesa*. São Paulo: Hedra, 2001.
- HOUAISS, Antônio; VILLAR, Mauro. *Dicionário Houaiss da Língua Portuguesa*. São Paulo: Objetiva, 2001.

- MACHADO FILHO, A. M. *O negro e o garimpo em Minas Gerais*. Rio de Janeiro: José Olympio, 1944.
- MENDONÇA, R. *A influência africana no português do Brasil*. Rio de Janeiro: Sauer, 1933.
- PETTER, M. M. T. Talvez sejam africanismos. *Estudos Lingüísticos XXIX – GEL – Grupo de estudos Lingüísticos do Estado de São Paulo*. v. 1, p. 713-718, 2000.
- _____. Palavras de origem africana nos dicionários Houaiss e Novo Aurélio. *Papia Revista Brasileira de Estudos Crioulos e Similares*, Brasília, v. 1, p. 107-114, 2003.
- PARADIS, Carole; LABEL, Caroline. Contrasts from segmental parameters settings in loanwords: core and periphery in Quebec French. *Proceedings of the MOT conference on contrasts in phonology*, n. 13, p. 75-94, 1994.
- RODRIGUES, Aryon. *Línguas brasileiras*. São Paulo: Loyola, 1986.
- SANDALO, F. Acento e sonoridade. *Reunião da Abralin*, 1999, Florianópolis/UFSC.
- SILVA, R. V. M. *O português arcaico: fonologia, morfologia e sintaxe*. São Paulo: Contexto, 2006.