

# DIME COMO REMATAS E DIREICHE ONDE TE ACENTÚO

VÍCTOR FRESCO BARBEITO  
Universidade de Vigo / Real Academia Galega  
rag.vfresco@gmail.com

## RESUMO

Cando se lles pretende asignar unha tónica ós neoloxismos chegados ó galego dende linguas onde a tonicidade non ten valor fonolóxico, o criterio de elección máis aconsellable é a propia tónica esperable en galego. Pero, cales son os parámetros que poden influír nesta escolla? A hipótese de partida nunha publicación previa do mesmo autor foi que dependía do derradeiro grafema da palabra. Neste artigo comprobaremos esta presuposición e tamén analizaremos se o número de sílabas ou a categoría gramatical poden influír como parámetros para determinar unha tónica prototípica. Para realizar este contraste, partimos dos datos do VOLGA e dun corpus de textos teatrais como exemplo de lingua en uso e aplicámoslles un algoritmo para a detección da tónica que se basea na acentuación gráfica e na terminación da forma escrita da palabra. De camiño, puidemos analizar tamén a distribución individual tanto da posición da tónica como dos tres supostos parámetros nestes dous conxuntos de datos.

## PALABRAS CHAVE

tónica, tónica esperable, número de sílabas, ortografía normativa, grafémica, terminación, adaptación de neoloxismos, categoría gramatical

## ABSTRACT

When trying to assign a stressed syllable to the neologisms arrived to Galician from languages where stressing have no phonological value, the most recommended criterium for the choice is to point towards the most expected stressed syllable for Galician. However, which are the parameters that may have an influence on this choice? The initial hypothesis in a previous article by the same author was that this depend on the last grapheme of the word. In this article we will check this presumption and will analyse whether the number of syllables or the grammatical category may count too as parameters to predict a prototypical stress syllable. To contrast this hypothesis we will use data from VOLGA and from a corpus made of theatre texts as an example of language in use, and we will apply an algorithm to detect the stressed syllable based on the graphic accents and the ending of the word written form. At the same time, we were able to analyse too the individual distributions for the grammatical category, the ending letter, the number of syllables or the position of the stressed syllable.

## KEY WORDS

stressed syllable, expected stress, number of syllables, standard orthography, graphemics, ending, adaptation of neologisms, grammatical category

### *1. Hipótese e antecedentes*

A orixe deste artigo é un problema que apareceu na aplicación dunha metodoloxía para a transcrición dos arabismos ó galego publicada no seu momento en *Cadernos de Lingua* (Fresco Barbeito 2007). Daquela houbo que propoñer que a tónica coa que se adaptarían estas palabras illadas, normalmente substantivos, dependería da terminación:

Debido ás dificultades que implica marcar unhas regras específicas para determinar a tónica [...] decidimos optar pola acentuación esoperable en galego e non basearnos no orixinal árabe. [...] Como hipótese de partida, a nosa intuición dinos que as palabras rematadas en vogal ou <s> deberían ser paroxítonas, e oxítonas tódalas demais. (Fresco Barbeito 2007:30)

Esta presuposición era moi arriscada, pero viña deducida da dobre característica na acentuación gráfica de evitala ó máximo posible e de depender da letra na que remata a palabra. Con esta investigación pretendemos confirmar esta hipótese e de paso contestar algunhas preguntas máis que os datos cos que traballamos permiten responder.

Para poder comprobar a nosa hipótese, precisamos coñecer dous parámetros de diferentes niveis: a letra final dunha palabra e a posición da tónica. Debe terse en conta que o primeiro é un parámetro gráfico, mentres que o segundo é un parámetro fonolóxico, polo que necesitamos ter un corpus coas tónicas marcadas (e o suficientemente grande para que nos permita extraer conclusións significativas). Como non contamos de momento con esta listaxe, consideramos que sería válido a conta das vogais dunha palabra. É dicir, buscaremos un resultado fonolóxico a través dunha análise gráfica. Somos conscientes de que hai marcos teóricos máis elaborados que permitirían un resultado máis acertado (Ohannesian 2004; Méndez 1997) e que a nosa simplificación pode levar a erros (como demostramos máis abaixo), pero tamén é certo que a necesidade de aplicalo automaticamente a corpus moi grandes é unha limitación para o algoritmo que busque a tónica, que cremos que a marxe de erro é asumible para os datos cos que traballamos e que aínda non existe unha norma de pronuncia que nos permita resolver casos graficamente ambiguos (como por exemplo, se certos encontros vocálicos son tautosilábicos ou heterosilábicos, cfr. Álvarez 2002:34).

Con todo, tres aproximacións anteriores poden sernos útiles e paga a pena telas en conta, cando menos para contrastar os seus resultados cos nosos: unha de Xosé María Álvarez Cáccamo en 1977, outra de Rodríguez Fer en 1984 e unha última de Xosé Luís Regueira en 2007.

Álvarez Cáccamo, aínda que centrado nas siglas, indícanos que xa en 1977 asumía por defecto que cando as vogais e consoantes se agrupan de acordo coas regras fonolóxicas “levan o acento intensivo na penúltima sílaba, agás [...] os monosílabos”, que “a tendencia das linguas románicas á acentuación paroxítona” e que “as palabras rematadas en consoante manifestan tendencia á acentuación aguda” (1997:210). O artigo de Rodríguez Fer e Botana énos útil

para comparar a distribución que obtivemos en final de palabra cos seus resultados para tódalas posicións (v. 3.3). O traballo de Regueira tamén restrinxe o seu enfoque, concretamente ós antropónimos, pero contrástaos con datos para o léxico máis frecuente do galego e en análises dunha obra súa anterior. Para este léxico común, afirma con carácter xeral:

- que o galego popular prefire “os nomes bisílabos, seguidos a distancia dos trisílabos” (Regueira 2007:241; dos gráficos 1 e 2 sácase: monosílabos: 10%; bisílabos: 50%; trisílabos: 30%; tetrasílabos: 10%);
- agudas 20%, graves 70%, esdrúxulas 5% (deducido dos gráficos 3 e 4 de Regueira 2007:243);
- “a configuración xeral das palabras do galego é de acentuación grave cando rematan en vogal e aguda cando acaban en semivogal ou consoante (non considerando neste grupo os plurais). Os demais casos (aguda rematada en vogal, esdrúxulas, etc.) son moi minoritarios.” (Regueira 2007:243)
- “son algo máis frecuentes as agudas de xénero masculino cás de xénero feminino (pénsese en parellas do tipo *rapaz/rapaza, autor/autora*.” (Regueira 2007:243; dos datos do gráfico 4 tírase: femininas agudas: 10%; masculinas agudas: 30%);
- “mentres que os femininos presentan un comportamento semellante ós nomes comúns. É dicir, mentres máis dun 40% dos nomes bisílabos de home son agudos, [...] nos de muller non chegan ó 20%” (Regueira 2007:244);

A terceira afirmación de Regueira para DG-PT conecta cos experimentos de repetición de Hochberg (Méndez 1997:264) e Gutiérrez Palma (2004) con nenos con lingua materna castelá, que se pode extrapolar ó galego por teren un sistema acentual case idéntico (Méndez 1997:255). A investigadora pedíalles a nenos pre-escolarizados que repetisen unhas palabras e os números seguintes amosan a porcentaxe de erro para cada configuración silábica (C: consoante; V: vogal) e tonicidade, o cal indica o grao de dificultade e antinaturalidade na lingua:

PT	4	3	2	1
CV.CV	-	-	7*	23
CV.CV.CV	-	20	13*	32
CV.CVC.CV	-	77	42*	75
CV.CV.CV.CV	50	30	14*	54
CV.CVC	-	-	50	22*
CV.CV.CVC	-	43	47	21*

Os valores máis pequenos (que marcamos con asterisco) indican a tonicidade prototípica para cada estrutura. Pode observarse que a explicación mencionada de Regueira serve como simplificación da táboa, na que tampouco se consigna ningunha esdrúxula como tonicidade prototípica de ningunha configuración.

Unha explicación da tonicidade en galego, aínda que diacrónica e polo tanto inviable para nós, é a de Ferreiro (1996:25): “Con relación ao acento, destaca a estabilidade da súa posición, manténdose, en xeral, a vogal tónica na evolución desde o latín ao galego -portugués. A posición do acento latino [...] [está] condicionada pola estrutura fonolóxica da palabra”.

## 2. *Corpus, metodoloxía e datos*

Como fonte dos datos, quixemos traballar con dúas orixes: unha é o VOLGA (2004), que nos fornece a categoría gramatical coma terceiro parámetro para controlar a súa posible influencia, pero que en contrapartida na nosa versión en base de datos se restrinxe inicialmente a 48941 lemas (e que reduciremos máis aínda), e a outra con formas flexionadas atopadas en textos reais. Nesta segunda fonte pretendemos buscar unha mostra de linguaxe non técnica, evitando os tecnicismos e estranxeirismos. Para iso, baseámonos nos textos teatrais modernos fornecidos pola Biblioteca Virtual Galega na súa páxina web da Universidade da Coruña (<http://bvg.udc.es/>). Restrinxímonos á produción do século XXI e, despois de eliminar dous

libros que seguían a norma ortográfica portuguesa (que pode interferir e mesmo inverter os resultados se se interpretan como escritos na norma galega) e outras cinco obras cun formato máis poético e narrativo ca oral, obtemos 25 publicacións. Tras un procesamento inicial conseguimos un primeiro corpus de 81.541 palabras moito máis variado có do VOLGA.

Con todo, quixemos tirar máis partido destas dúas orixes, polo que creamos oito distribucións distintas. Para denominalas utilizamos a convención de comezar coa orixe (T: teatro; V: VOLGA) e seguir coas súas características: se tiñamos en conta (CONR) ou non (SENR) as palabras repetidas, se conservamos (CONM) ou non (SENM) os monosílabos e se separamos os posibles estranxeirismos (STRAN) das palabras máis adaptadas. Polo tanto creamos as seguintes distribucións:

- T-CONR-CONM
- T-CONR-SENM
- T-SENR-CONM
- T-SENR-SENM
- T-STRAN-SENR-CONM
- V-SENR-CONM
- V-SENR-SENM
- V-STRAN-SENR-CONM

Para poder facer comparacións válidas entre estas distribucións de tan diferentes tamaños, utilizamos os valores en porcentaxes sobre o total.

Para identificar os estranxeirismos utilizamos varios sistemas:

- Aproveitar as marcas do VOLGA, pois nas súas notas aparecen ás veces a orixe e mesmo o étimo ou a pronuncia, aínda que non sistematicamente. Descartamos porén os préstamos coas marcas de “lat”, “gr” e “cast”. Estas mesmas palabras (aínda que só as non flexionadas) tamén se marcaron como estranxeirismos no corpus teatral.

- Localizar as palabras cunha secuencia gráfica determinada. Para iso identificamos como estranxeirismos tódalas palabras que contivesen os grafemas J, K (e KH), W, Y e moitas cunha dobre consoante (que non fose -CC-, -LL- ou -RR-; é dicir, con -BB-, -DD-, -FF-, -GG-, -MM-, -PP-, -SS-, -TT- e -ZZ-), con dobre vogal (-EE-: *feedback, jeep, neerlandés*; -OO-: *zoom, footing*; -UU-: *continuum, weltanschauung*), determinados dígrafos (-GE-: *geyser, hegelianismo, hegeliano, roentgen, röntgen, sui generis*; -GI-: *adagio*; -GH-: *copyright, gheada, malpighiáceas, oghámico, spaghetti*; -QUA-: *exequátur, quásar, squash*; -QUO-: *quórum, sequoia*; -SH-: *flash, leishmania, leishmaniose, shakesperiano, shériff, sherpa, smash, squash, washingtoniano, yiddish*) ou trígrafos (-SCH-: *kirsch, nietschiano, scherzando, scherzo, weltanschauung*) ou uns grafemas determinados en posición inicial (S- + consoante; I- + vogal) ou final (-B, -C, -D, -F, -G, moitos en -H, -K, -M, -P, -T, -V, -W, -X, -Y, -Z; incluídas como estranxeirismos debido a ter unha destas terminacións foron tamén bastantes interxeccións: *clac, clic, tac, tic, tictac; paf, pataplaf, pof, pluf, uf; ab, bah, eh, ob, ouh; hum, pim, pam, pum; up; chist; ux*).
- Algúns topónimos (*Aranjuez, Gibraltar, Himalaya*, e xentilicios: *frankfurtés*) e antropónimos (*Cage, Wayne, Zalassokrátés*) que presentan estes cadeas gráfica e mesmo outros nomes propios detectados sen elas pero cunha forma pouco ou mal adaptada ó galego (*Banánia, Caimánia, Coimbra [sic], Iónico [sic, por Xónico], Toulousse, Madrid, Führer*).
- Os exemplos que dá Rodríguez Río (1998). Con todo, aínda persisten algunhas grallas na aplicación das regras acentuais a palabras do VOLGA (*angstrom, geyser, yiddish*), mentres que outras algunhas están suprimidas (*sponsor*), a maioría corrixiadas (*macádam, máxwell, módem, stárter*) e algunhas nin sequera detectadas antes (*vademecum, ibidem, passim; napalm?*). A maiores, aínda que o VOLGA non é máis ca un repertorio de formas e non un dicionario de dúbidas, bótase en falta unha explicación máis detallada e mesmo unha clasificación da información dada nas notas e do formato con que aparece, pois ás veces márcase a lingua de orixe para dar simplemente a etimoloxía, a forma ou pronuncia orixinal, o topónimo

dun xentilicio, a flexión, defínilo coma estranxeirismo non adaptado ó galego (co cal debe grafarse en cursiva nun texto) ou nada en absoluto, todo o cal se fai asistematicamente con comiñas, corchetes ou tipos de letra que non permiten moito xogo á hora de representar cadeas fónicas.

- Os plurais simples das anteriores: *bares, jockeys, pizzas, restaurantes, tics*.

Con todo, e a pesar de que será interesante observar o seu comportamento diferencial respecto ó léxico común, somos conscientes de que agrupar estranxeirismos de distintas orixes é unha gran simplificación, pois para unha análise da tonicidade, mentres que os galicismos son orixinariamente sempre oxítonos, noutras linguas hai maior variación, e isto ademais sen considerar a morfoloxía, a época e o grao de implantación no galego.

O procesamento inicial no corpus teatral consistiu na eliminación de fragmentos en castelán (351 palabras da obra *Falando con balas*), 60 números romanos, palabras cortadas, signos de puntuación e ortotipografía (marcas introdutoras de personaxe), mentres que no VOLGA eliminamos cadeas multiverbais (342 locucións), palabras compostas con guión (630 casos), palabras incompletas (602 prefixos), palabras agramaticais (149 símbolos, siglas e abreviacións) e 1.406 palabras marcadas coma incorrectas. Tras estas eliminacións e asignacións coma estranxeirismos, obtivemos que as distribucións contiñan a seguinte cantidade de datos:

T-CONR-CONM	78.311
T-CONR-SENM	49.192
T-SENR-CONM	13.330
T-SENR-SENM	13.043
T-STRAN-SENR-CONM	320
V-SENR-CONM	45.725
V-SENR-SENM	45.477
V-STRAN-SENR-CONM	693



Debemos explicar polo tanto cal foi o procedemento na análise da cadea gráfica que seguimos para detectar a tónica. O primeiro paso foi a eliminación de tódalas consoantes. Houbo que ter coidado especial en eliminar, manualmente, o -U- do grupo -GU- só cando fose mudo (e na inmensa maioría dos -QU-) e en desenlear a lectura consonántica ou vocálica de -Y-. Respecto a -X-, a súa lectura monofonemática ou bifonemática non afecta á súa interpretación como terminación. Despois desta eliminación, atopamos dúas situacións nesta palabra sen consoantes:

- A palabra xa ten unha tónica gráfica, polo que xa queda automaticamente detectada.
- A palabra non presenta tónica gráfica, polo que debe deducirse en función da terminación da palabra orixinal e segundo estipulan as NOMIG (2005). Neste caso:
  - Se remata en vogal, -N ou -S, a palabra é paroxítona.
  - Se remata no resto de consoantes, a palabra é oxítona.

Cómpre comentar que esta análise non debe aplicarse ós monosílabos, pois non teñen unha penúltima vogal á cal asignarlle a tónica, e polo tanto serán oxítonas.

Como comentamos antes, tanto as palabras con tónica gráfica como as que non a teñen presentan o problema de que a simple conta das vogais pode levar a deducir unha tónica incorrecta. Isto débese a que a eliminación, sen máis, das consoantes pode levar a datos iniciais incorrectos para o algoritmo, pois unha vogal non ten por qué corresponder a unha sílaba. Mesmo así, ás veces o algoritmo produce resultados correctos. Sen pretender ser sistemáticos, uns exemplos destas situacións serían os seguintes:

- hiato final: *velaí* e *badía*, que darían correctamente a súa tónica por estar marcada graficamente;
- hiato non final: *triada* ou *sáinte* quedarían respectivamente “íaa” e “aíe”, polo que se deduciría correctamente que son esdrúxula e grave;
- ditongo átono final, chamado semiesdrúxula: *valencia* daría como resultado sen consoantes “aeia”, sen tónica gráfica,

polo que os datos de partida do algoritmo serían incorrectos (tetrasílaba) pero a conclusión como paroxítona sería correcta; *reis* daría, mediante o procesamento de “ei” un falso bisílabo paroxítono;

- ditongo átono non final: *xeito* ou *reitor* teñen a mesma secuencia de vogais (“eio”) mentres que a tónica está en diferente posición, pero a terminación diferente (antes de quitarlle as consoantes) permite descubri-la;
- ditongo tónico final: *cantei*, con secuencia de vogais “aei” que indica incorrectamente tres sílabas e implica incorrectamente unha paroxítona (por non contar cun acento gráfico e estar rematada en -I);
- ditongo tónico non final: *fendo* é paroxítona (e a secuencia das súas vogais “euo” implica incorrectamente unha posición correcta);
- pronuncias variables (ditongo ou hiato) segundo falantes: *ciencia* pode pronunciarse trisílaba ou quadrisílaba (Álvarez 2002:34), pero en todo caso a tónica deducida (coa serie de vogais “ieía”, pola que se lle asignaría incorrectamente 4 sílabas) é correcta (paroxítona);
- acentuación diacrítica: no caso de *máis* deduciríase incorrectamente que é paroxítona; no caso de *só* ou *présa*, os resultados son correctos;

Unha mellora do algoritmo sería o mantemento da posición ocupada por calquera consoante para poder detectar os ditongos, e imporlles máis condicións ás deducións, pero esta complicación non semella rendible.

Os exemplos amosan que o algoritmo funciona moitas veces, mesmo con datos de partida (número de sílabas) incorrectos. Ademais, o erro que se produce mediante esta simplificación é unidireccional: unha palabra con tónica na posición 3 (esdrúxula) pode acabar identificada coma tónica en 3 ou máis por non considerarse a conxunción de vogais postónicas en ditongos, pero nunca se lle poderá asignar unha tónica menor ca 3 porque en galego “o núcleo da sílaba é sempre unha vocal” (Álvarez 2002:42). Para verificar a fiabilidade do algoritmo realizamos unha comprobación manual dos

resultados para un grupo aleatorio de palabras consecutivas do corpus teatral (onde previamos máis erros de identificación debido a que a morfoloxía flexionada é máis irregular, presuposición que o grupo aleatorio demostrou que era correcta), que pertenceron ó libro *Matanza* de Roberto Salgueiro. Nun total de 250 palabras houbo 10 erros, correspondentes todos a palabras rematadas en ditongo oxítono identificadas como paroxítonas: *ou, seu, vai, hai, máis, demáis* [sic], *estou* e *tratarei*. Asumindo este erro (4%) como aceptable, parécenos que terá pouca influencia á hora de confirmar a nosa hipótese.

Os parámetros nos que se basea a hipótese son dous: a posición da tónica e a terminación, en realidade o derradeiro grafema. Para analizar a localización da vogal tónica simplificamos os resultados mediante valores numéricos: unha oxítona está na posición 1, unha grave na 2, unha proparoxítona na 3, unha sobresdrúxula na 4, etc. Grazas a isto puidemos calcular medias.

Dado que o algoritmo procesa as palabras quitándolles tódalas consoantes tanto antes coma despois da tónica, en realidade tamén pode fornecer como resultado colateral o número de sílabas. Este valor é moito máis inexacto có da posición da tónica, pois non contamos con outro valor corrector (como o acento gráfico ou a terminación cando se buscaba a posición) e o erro pode estenderse acumulativamente ó longo da cadea de vogais. Con todo, o erro no cálculo da cantidade de sílabas tamén é unidireccional como o da posición da tónica, polos mesmos motivos. Terémolo en conta por se permite sacar algunha conclusión, tamén grazas ó ocasional uso dos valores medios.

Un cuarto valor co que se pode contar só no caso dos datos obtidos do VOLGA é a categoría gramatical. No VOLGA aparecen as palabras marcadas con tódalas súas posibles categorías, polo que existen 203 combinacións: por exemplo, unha palabra pode ser “a f”, outra “af f” e outra “f af”. Independentemente da pertinencia na orde da caracterización ou da súa combinación cunhas ou outras categorías gramaticais, hai 32 marcacións básicas para categorías. Con todo, consideramos que son irrelevantes (salvo para comparar con algunhas conclusións de Regueira) outras especificacións máis profundas coma masculino, feminino, transitivo ou pronominal, polo que decidimos agrupar as máis importantes en seis grupos:

- substantivo (S): f, fpl, m, mpl, s, ls;
- adxectivo con uso coma substantivo (SA): af, am, as;
- adxectivo (A): a, la;
- adxectivo con uso coma verbo (AV): pt;
- verbo (V): t, i, pr, aux, cop;
- adverbio (B): av, lav;
- tras simplificar os símbolos nun só grupo (abrev, sb), o resto das categorías queda coma no orixinal: numeral, pronome, artigo, contracción, conxunción, interxección, preposición e prefixo;

Con todo, debe terse en conta que cada categoría gramatical agrupada nova crea un novo rexistro. É dicir, un verbo intransitivo, pronominal ou copulativo (*estar*) contará unha vez, pero un verbo ou substantivo (*poder*) contará dúas veces, por iso a suma total de CX é maior cá da distribución doutros parámetros para o VOLGA:

CX + V-SENR-CONM	52.682
CX + V-SENR-SENM	52.405
CX + V-STRAN-SENR-CONM	763

Para mencionar estes catro parámetros e aforrar espazo usamos as seguintes abreviaturas:

- PT: posición da tónica
- DG: derradeiro grafema
- NS: número de sílabas
- CX: categoría gramatical xeral do VOLGA

### 3. Distribucións individuais

#### 3.1. Distribución da categoría gramatical

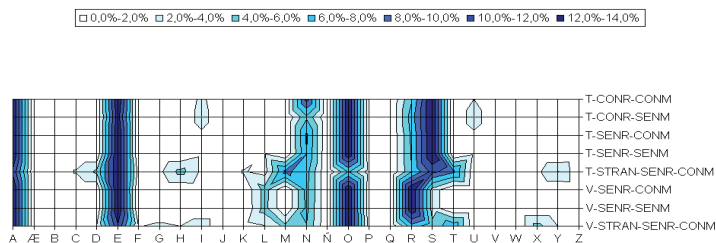
Os resultados para as principais categorías gramaticais (V-SENR-CONM) son:

substantivo (S)	28.621
adxectivo (A)	12.977
verbo (V)	8.934
adverbio (B)	1.857
participio (AV)	573
adxectivo2 (SA)	117

Isto amosa unha maioría de substantivos, adxectivos, verbos e adverbios, esperable debido a que son as categorías máis abertas de todo o rango gramatical. Cando se desagregan, os resultados maiores de mil son “f”: 12.087; “m”: 10.445; “a”: 8.689; “t”: 3.069; “a s”: 2.750; “av”: 1.800; “t pr”: 1.325; “s”: 1.210; “i”: 1.011, amosando outra vez a superioridade na frecuencia de substantivos e adxectivos.

### 3.2. Distribución do derradeiro grafema

As posibilidades de terminación dunha palabra varían relativamente pouco en función da súa orixe (GRÁFICA 1: derradeiro grafema). Os valores máis comúns en tódalas distribucións son os das vogais -A, -E, -O e das consoantes -N e -R ou -S.



GRÁFICA 1: derradeiro grafema

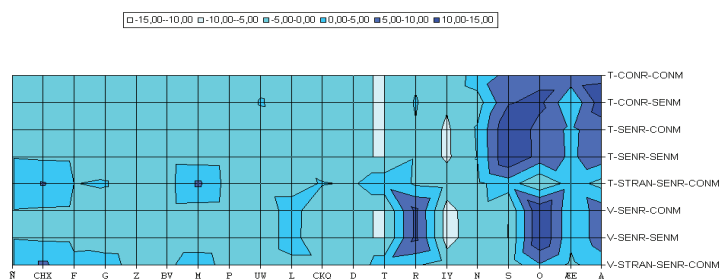
A distribución dos lemas no VOLGA sen incluír os monosílabos mostra unha tendencia cara á terminación en vogal non pechada e en líquida. Cando se inclúen os monosílabos, a orde nos resultados non presenta variación para a metade das terminacións con máis frecuencia. A especificidade do VOLGA de expresar os verbos mediante o infinitivo é a causa de que presente unha maior presenza de -R (en cifras aproximadas, das once mil palabras en -R, nove mil son verbos). A maior presenza de valores de -L nos datos do VOLGA débese ós adxectivos: mentres que nos textos teatrais só aparece ou ben a forma en -BLE ou en -BEL para os adxectivos, no VOLGA aparecen as dúas formas máis de 900 veces cada unha (de entre 12.000 adxectivos).

O que si é máis frecuente nos corpus de fala flexionada son as terminacións -N, -S e -A, debido a que as palabras flexionadas aumentan as súas aparicións por seren as marcas habituais de plural verbal e plural e feminino nominal respectivamente, que no VOLGA aparecen baixo o lema en infinitivo ou masculino singular. As terminacións -I e -U tamén aumentan a máis dun 2% no corpus teatral sobre todo en T-COND-CONM e T-COND-SENM, que tamén é debido á conjugación verbal, neste caso do pretérito.

As distribucións de estranxeirismos T-STRAN-SENR-CONM e V-STRAN-SENR-CONM teñen a particularidade de presentar terminacións atípicas nas outras distribucións: -T, -Y, -K, -I, -H, -G ou -F. Un caso especial é o -Æ presente na expresión DRAMATIS PERSONÆ da obra de Nacho Taibo. Aparecen tamén un -Ê, dous -Ô e dous Â, que se incluíron baixo as vogais sen ningún tipo de acento. Con todo, non sorprende que as terminacións dos estranxeirismos coincidan coas das outras distribucións, pois a lingua ás veces “recorre a vogais de apoio no proceso de integración, [...] (*bucle, cobalto, cheque, casete, envite, iate, ponche, tanque*) [...] [mentres que hai] outros semellantes en que non o fai (*debut, complot, fagot, boicot, sándwich, iceberg, slogan, smóking, snob, sprint, starter, stock...*).” (Rodríguez Ríu 1998:305).

### 3.3. Distribución do derradeiro fonema

Para poder contrastar fonemas e grafemas finais cómpre facer varios grupos. Nos fonemas, sumamos os diferentes valores das vibrantes nun só valor. Por cuestións prácticas, tamén xuntamos os valores dos fonemas palatais fricativo e africado, pois moitas veces veñen representados. Nos grafemas, agrupamos aqueles que se len igual en posición final de palabra: -B + -V, -U + -W, -I + -Y, -Æ + -E e -C + -K + -Q. Para os fonemas /b/ e /p/ tivemos que estimar os valores como 2,80 e 2,35, pois non se explicitan na fonte orixinal.

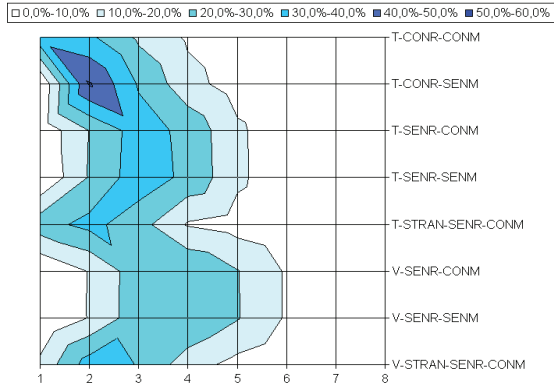


GRÁFICA 2: % frecuencia fonémica final - % frecuencia fonémica xeral

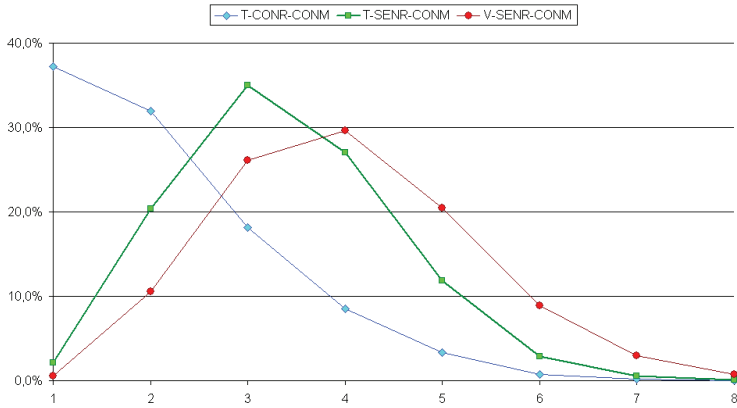
Como se pode observar na gráfica (GRÁFICA 2: frecuencia fonémica final / frecuencia fonémica xeral), unha boa parte dos fonemas presenta unha frecuencia en posición final absoluta semellante ou pouco menor á xeral que ten para tódalas posicións da palabra. Por iso é máis rechamante que aumente a frecuencia das vogais /a/ e a suma das vogais medias velares (-O), mentres que diminúe a das suma das vogais medias palatais (-E). A frecuencia de /r/ final aumenta no VOLGA debido a que os verbos teñen todos esta terminación, mentres que a de /l/ aumenta en tódalas fontes. Sorprende tamén que a maior baixada sexa no fonema pechado palatal (-I), pois tamén está presente en terminacións verbais, pero tódolos comportamentos cambian cando se consideran os estranxeirismos: baixan os fonemas máis habituais (salvo -O no VOLGA) mentres que soben outros fonemas máis escasos como /m/, /g/ ou /f/. O caso dos fonemas

palatais agrupados comparados cos grafemas -CH + -X é un dato falseado porque tamén se incluíron os valores de -H.

### 3.4. Distribución do número de sílabas



GRÁFICA 3: número de sílabas



GRÁFICA 4: número de sílabas con monosílabos



A cantidade de sílabas por palabra ten unha variación maior cá anterior (GRÁFICA 3: número de sílabas). A presenza de monosílabos dispárase na distribución con repetición (T-CONR-CONM). Isto débese á enorme frecuencia deles na fala común. Unha comparación específica entre as distribucións con monosílabos (GRÁFICA 4: número de sílabas con monosílabos) permite observar que a distribución acampanada do VOLGA baixa o seu máximo no corpus teatral sen repeticións. Pero cando se considera a frecuencia de uso dos monosílabos, a distribución gaussiana convértese nunha hiperbólica, onde a frecuencia é maior cantas menos sílabas ten a palabra. Loxicamente, as distribucións sen repetición presentan medias máis altas, de 3 e 4 e mesmo 5 sílabas, en realidade porque a presenza de monosílabos é moi frecuente nas que contan con repetición. Isto vén amosado tamén polos seguintes valores medios:

T-CONR-CONM	2,12
T-CONR-SENM	2,78
T-SENR-CONM	3,38
T-SENR-SENM	3,43
T-STRAN-SENR-CONM	2,46
V-SENR-CONM	4,02
V-SENR-SENM	4,04
V-STRAN-SENR-CONM	2,73

Con todo, o VOLGA presenta unha frecuencia de pentasílabas maior ca nas distribucións teatrais sen repetición: un repaso dos datos amosa que se trata principalmente de cultismos, adverbios rematados en -MENTE e outros derivados. A amálgama dos pronomes átonos co verbo nos textos teatrais non compensa estes lemas. Este dato será maior có real debido ó comentado anteriormente sobre o erro de asumir o resultado da eliminación de consoantes coma unha conta das sílabas. Se se eliminan os 480 monosílabos (e outros datos con circunstancias adicionais descritas na casuística da categoría) evidentemente aumenta a media, pero son relativamente tan poucos que o fai só en dúas centésimas. Como

curiosidade, atopamos que as palabras con maior número de sílabas adoitan presentar na súa formación as secuencias -MENTE, -DADE e sobre todo -CIÓN (mesmo en interior de palabra: *inconstitucionalidade*).

Os datos do teatro amosan unha presenza de palabras de 3 sílabas maior cá dos lemas do VOLGA, que probablemente sexa debida á enorme variación flexional de verbos, nomes e adxectivos.

Curiosamente, as dúas distribucións de estranxeirismos presentan unha media máis baixa, unha maioría de bisílabos, unha maior presenza de monosílabos e son moi similares entre elas (con valores xerais de NS un pouco máis elevados no caso da do VOLGA).

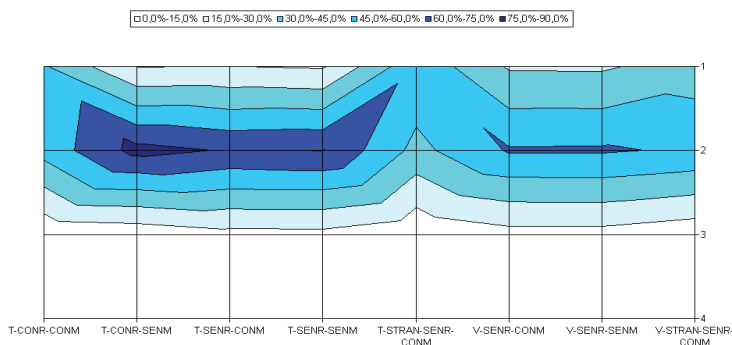
De tódalas nosas distribucións, as máis semellantes ós cálculos de Regueira (2007:242) sobre o número de sílabas para nomes comúns (10%-50%-30%-10%, obviamente con monosílabos) son as dúas de teatro con repeticións (T-CONR-CONM: 35%-30%-15%-10%-5%; T-CONR-SENM: 0%-50%-30%-15%-5%) e as de estranxeirismos (T-STRAN-SENR-CONM: 25%-35%-25%-10%-5%; V-STRAN-SENR-CONM: 15%-35%-30%-15%-5%). Polo tanto, poderíamos deducir que os seus datos contemplaban as repeticións, pois o resto das distribucións prefiren os tetrasílabos (as do VOLGA) e os trisílabos. Con todo, a máis probable é a T-SENR-CONM (2%-20%-35%-27%-11%), pois debemos lembrar que o erro no algoritmo é unidireccional e crecente.

### 3.5. Distribución da posición da tónica

A situación da tónica é bastante constante entre distribucións e desigual para os distintos valores de PT. Como se pode observar na figura correspondente (GRÁFICA 5: posición da tónica), salvando cando se consideran os monosílabos e as repeticións ou os estranxeirismos, máis do 60% son oxítonas. Os valores medios tenden principalmente á posición paroxítona e son menores nas distribucións con repetición, con monosílabos ou de estranxeirismos:

T-CONR-CONM	1,57
T-CONR-SENM	1,91

T-SENR-CONM	1,96
T-SENR-SENM	1,98
T-STRAN-SENR-CONM	1,47
V-SENR-CONM	1,82
V-SENR-SENM	1,82
V-STRAN-SENR-CONM	1,68



GRÁFICA 5: posición da tónica

Os valores máximos para as graves son máis marcados nos textos teatrais, debido ás repeticións reais e ás repeticións encubertas, é dicir, ás diferentes formas flexionadas que no VOLGA se simplifican nun só lema pero que no corpus teatral teñen cada unha a súa aparición. Os monosílabos son tan frecuentes que, cando se consideran as súas repeticións, desprazan a distribución ata case igualar as oxítonas (45%) e as paroxítonas, (50%) reducindo tamén a presenza de proparoxítonas (5%). Loxicamente, unha omisión das abundantísimas oxítonas monosilábicas reduce a súa presenza ó 15% do total, mentres aumenta a porcentaxe das paroxítonas a máis do 70%.

Nos estranxeirismos tamén se igualan estas dúas acentuacións, para o cal pode haber ata tres explicacións: que moitos dos estranxeirismos sexan monosílabos (como xa vimos), que nunha das linguas orixinais dos estranxeirismos (o francés) tódalas palabras

sexan agudas ou que hai moita frecuencia de estranxeirismos rematados en consoante, polo que os lectores galegos que len a palabra sen acento gráfico por primeira tenden a pronunciala coma oxítona.

As nosas distribucións permiten deducir outra vez que Regueira traballou só coas distintas formas flexionadas, sen considerar as frecuencias de repetición, pois na gráfica 5 podemos ver que as que máis se aproximan ás porcentaxes do seu gráfico 3 (Regueira 2007:243) son T-SENR-CONM e T-SENR-SENM.

#### 4. *Distribucións cruzadas*

A combinación dos catro parámetros considerados anteriormente produce seis novos grupos de datos con dous parámetros (CX-DG, CX-NS, CX-PT, DG-NS, DG-PT e NS-PT).

Para calquera destas distribucións, sempre hai dúas porcentaxes: a que se obtén de dividir cada resultado entre a suma de tódolos valores cos que coincide en valor no primeiro parámetro ou no segundo. De cada porcentaxe dedúcense tendencias distintas: por exemplo, en CX-DG, o dato para “verbo e terminación en R” pode dividirse entre o total de verbos ou entre o total de R: no primeiro caso, dará o 100%, pois tódolos verbos rematan en R, pero no segundo dará menos, os verbos en R son só unha parte de tódalas palabras rematadas en R. Con todo, os resultados dunha delas serán case sempre moito máis interesantes.

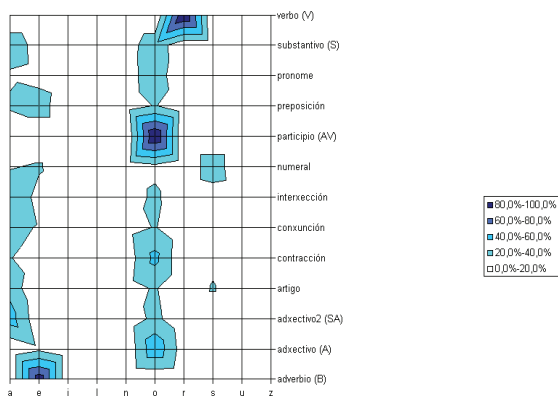
##### 4.1. *Distribución CX-DG*

Na primeira combinación CX-DG (GRÁFICA 6: DG/CX en V-SENR-CONM) poden destacarse os seguintes resultados esperables (e obvios algúns deles):

- a exclusividade de -R nos verbos
- en contraste, a presenza de substantivos con tódalas terminacións posibles (e case tamén para adxectivos)
- a case exclusividade de -O nos participios (salvo para *preñe*)
- a predominancia de -O nos adxectivos, e menos en substan-

tivos, pronomes e contraccións (31 casos dun total de 67), pero de -A nos adxectivos-substantivos e artigos (3 casos: *a, unba, la*)

- a maioría de vogal non pechada nos substantivos
- a predominancia de -E nos adverbios, dos cales só 26 non rematan en -MENTE
- a relativa abundancia de -S nos numerais (15 casos dun total de 44)
- a predominancia dos substantivos en tódalas terminacións excepto -F, -L e -R, onde son maioría respectivamente as interxeccións (con 5 casos: *paf, pataplaf, plaf, puf, uf*), os adxectivos e os verbos. Tamén en -H son abundantes as interxeccións (3 casos dun total de 16: *ab, eb, ob*) e en -O os adxectivos (debido á forma de enuncialos en masculino singular).



GRÁFICA 6: DG/CX en V-SENR-COMM

Estes datos quedan esquematizados na seguinte táboa:

S	todas
SA	-A, -O
A	-O, -L, -R, -A
AV	-O (-E)
V	-R
B	-E (-O, -A, -S)

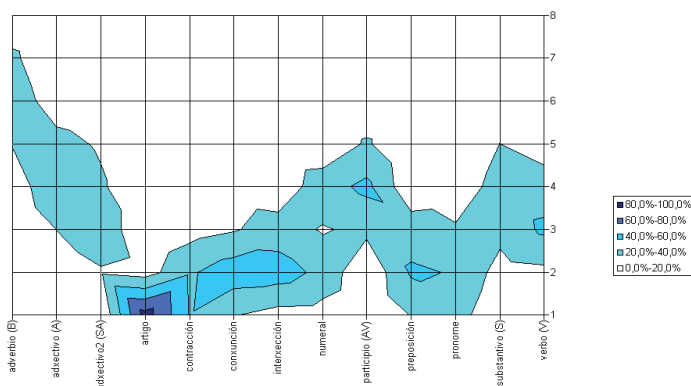
Estes resultados mudan pouco cando se omiten os monosílabos, e os cambios afectan principalmente a artigos, interxeccións, contraccións e pronomes. Neste caso, coa desaparición das interxeccións monosilábicas, os substantivos convértense na categoría gramatical maioritaria das palabras en -H.

En cambio, cando se consideran os estranxeirismos, a situación modifícase radicalmente: a interxección pode rematar unicamente en -O (1 caso: *alto*), terminación tamén maioritaria para o adverbio (*piano*, *pianísimo*, *scherzando*, *adagio*, *allegro* e *allegretto* fronte a *andante* e *passim*), e o verbo segue a rematar exclusivamente en R. O substantivo xa é maioritario en tódalas terminacións salvo en -X, terminación na que cede parte do seu posto maioritario ós adxectivos (5 casos de 37: *unisex*, *dúplex*, *múltiplex*, *réflex*, *súx*).

#### 4.2. Distribución CX-NS

Na relación entre a cantidade de sílabas e a categoría gramatical (GRÁFICA 7: NS/CX en V-SEN-R-CONM) pode verse que as proporcións máis altas son as seguintes:

- nos artigos, os monosílabos;
- nas conxuncións e nas preposicións, as bisílabas;
- nos verbos, os trisílabos;
- nos participios, os tetrasílabos;



GRÁFICA 7: NS/CX en V-SEN-R-CONM

Cómpre notar que a maioría dos adverbios son hexasílabos, que resulta ser dúas sílabas máis cós adxectivos puros e tres máis cós adxectivos-substantivos. Isto indicaría que moitos se construírían con -MENTE (o cal é certo: de 1857 adverbios, só 180 non rematan así) e que preferentemente se constrúen con adxectivos que non son substantivos. Isto parece ser certo pois certos adxectivos con valor case substantivo semellan impedir o adverbio: *bisectrizmente, pluscuamperfectamente, alisamente, esterlinamente, maionesamente, claudiamente, encintamente*, etc.

Cando as porcentaxes se fan sobre o número de sílabas, os substantivos teñen máis dun 40% en tódolos valores, pero hai máis dun 20% en adxectivos (para 4, 5, 6 e 11 sílabas; este último caso débese a unha gralla do VOLGA: *lúcidamente*) e en adverbios (para 7, 8, 9 e 10 sílabas). Cómpre lembrar que estes valores de NS son sempre maiores ou iguais ós reais.

Cando se eliminan os monosílabos, non hai cambios significativos salvo nas que teñen medias máis pequenas: preposición, conxunción, interxección, pronomes, numerais e contraccións. Si hai variación cando se trata dos estranxeirismos pois:

- non presenta conxuncións, numerais nin participios;
- os máximos para as oxítonas de preposición, interxección e contracción son máis acentuados;
- sobe a media dos adxectivos, cun máximo nas 4 sílabas;
- baixa a media dos adverbios, de 6 a 3, debido á presenza de poucos adverbios entre os estranxeirismos;

Nas porcentaxes de cada categoría gramatical para cada NS en particular, os resultados para estranxeirismos son idénticos ós das distribucións anteriores.

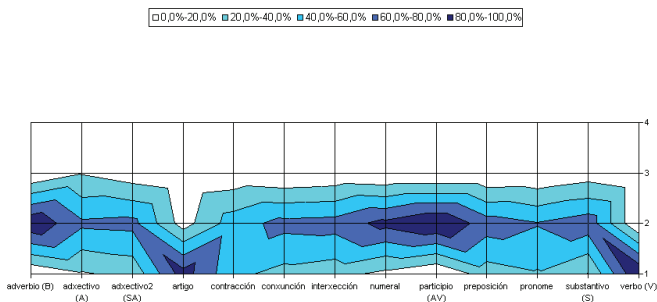
Observamos tamén o máximo posible de sílabas para cada categoría e vemos que NS é maior canto máis aberta é CX. Neste aspecto coincide en certa maneira coa formulación da Teoría cognitiva recompilada por Lakoff (1987:32-33 citando a Brown e 1987:46 citando a Rosch) respecto a que as palabras de nivel básico son máis breves e simples, máis frecuentes (que veremos no apartado do corpus

teatral), máis comprensibles e lembrables neuronalmente e máis adquiribles polos meniños, entendendo como palabras de nivel lingüístico básico as que fornecen os significados máis gramaticais.

substantivo (S)	11
adxectivo (A)	11
adverbio (B)	9
verbo (V)	8
participio (AV)	7
adxectivo2 (SA)	6

### 4.3. Distribución CX-PT

Se a combinación é CX-PT (GRÁFICA 8: PT/CX en V-SENR-CONM), observamos unha continuación da tendencia xeral de maioría paroxítona, máis pronunciada en substantivos e sobre todo en adverbios e participios. As oxítonas van aparecendo con proporción crecente dende as preposicións (28%), as interxeccións, as conxuncións, os pronomes, as contraccións (44%) ata os artigos (8 de 10) e anularse sistematicamente nos verbos, pero todos estes valores volven por debaixo do 20% cando se eliminan os monosílabos. Cando a distribución se calcula en función de NS, as probabilidades son máximas para verbos en agudas, para substantivos en graves e esdrúxulas e para adxectivos en esdrúxulas e sobresdrúxulas.



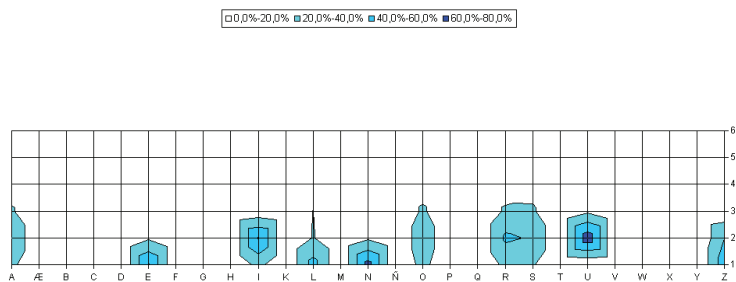
GRÁFICA 8: PT/CX en V-SENR-CONM



Os estranxeirismos (para os que só hai datos con substantivos, adxectivos, adverbios, verbos, interxeccións e un único pronome, *idem*) continúan estas tendencias oxítonas nas frecuencias para cada CX (salvo para o pronome).

Considerar o parámetro CG separando masculino e feminino permítenos confirmar ós cálculos de Regueira (2007:243) sobre que son “máis frecuentes as agudas de xénero masculino cás de xénero feminino”: os resultados tirados dos datos do VOLGA varían dende o 17% de agudas nos femininos e o 16% nos masculinos en V-SENR-CONM e V-SENR-SENM ata o 2% no feminino e o 44% nos masculinos en V-STRAN-SENR-CONM. Cando se restrinxen estes datos ós bisílabos, Regueira (2007:244) calcula que hai máis graves (80%) entre os bisílabos femininos (e polo tanto agudos son o 20%) ca entre os bisílabos masculinos (60%). Os nosos resultados para masculinos dan a mesma porcentaxe nas tres distribucións do VOLGA, pero para femininos obtemos resultados máis altos aínda: un 95% de graves en V-SENR-CONM e V-SENR-SENM e o 100% en V-STRAN-SENR-CONM.

4



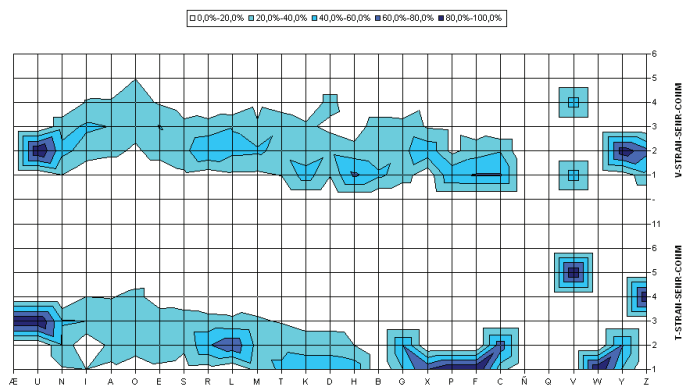
GRÁFICA 9: NS/DG en T-CONR-CONM

Respecto ó número de sílabas que pode observarse para cada terminación (GRÁFICA 9: NS/DG en T-CONR-CONM), son -E, -L, -N e -Z nas que predominan os monosílabos e -A, -I, -O, -R,

-S e sobre todo -U as terminacións nas que son máis habituais os bisílabos. A exclusión dos monosílabos produce que as frecuencias se repartan para as palabras de 2, 4 e principalmente 3 sílabas en todas estas terminacións. Cando as distribucións son as do VOLGA, -I, -U e -Z presentan frecuencias máximas para 2 e 3 sílabas; -A, -L, -O, -R e -S para 3, 4 e 5; -E para 4 e 5; e -N para 5 sílabas.

Se as porcentaxes se fan sobre NS, constatamos que -E é maioritario nas palabras de entre 6 a 9 sílabas, mentres que nas menores é -A e -O se se consideran as repeticións ou -A e -S se non. No VOLGA, -O e -R é maioritario entre 2 e 6 sílabas, o -E segue a selo entre 6 e 10 e -N e -O nas 11 sílabas.

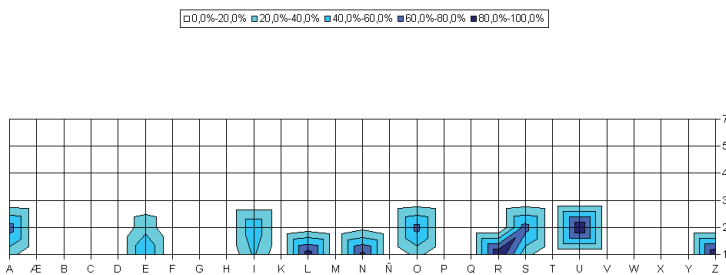
A situación nos estranxeirismos é moito máis heteroxénea, sobre todo no corpus teatral. Cada fonte ten uns resultados distintos, aínda que -H, -F, -P e -K semellan tender ó monosilabismo, -C, -G, -L, -R e -Y ó bisilabismo e -N, -O, -E, -M, -A e -S ó trisilabismo. Detectamos tamén a coincidencia de que, cando se ordenan as terminacións do corpus teatral pola frecuencia de trisílabos, a mesma orde fornece unha distribución moi similar nos datos do VOLGA (GRÁFICA 10: NS/DG en T-STRAN-SENR-CONM e V-STRAN-SENR-CONM).



GRÁFICA 10: NS/DG en T-STRAN-SENR-CONM e V-STRAN-SENR-CONM

#### 4.5. Distribución DG-PT

A relación entre DG e PT é a hipótese de partida deste traballo e o primeiro que observamos é que no VOLGA os resultados están máis espallados. Dado que a variación non é grande no corpus teatral (GRÁFICA 11: PT/DG en T-CONR-CONM), tanto se se consideran os monosílabos ou as repeticións coma se non, a concentración de resultados nunha tonicidade determinada para cada terminación nas formas flexionadas débese á redundancia provocada pola frecuencia nas terminacións, non nas palabras.



GRÁFICA 11: PT/DG en T-CONR-CONM

Os datos en porcentaxe no corpus teatral cando se refiren ó total de cada terminación amosan que a terminación -U é case exclusivamente paroxítona (98%) e en menor medida -A (69%), -S (69%), -O (67%) e -I (57%). Cando non se consideran as repeticións ou os monosílabos, estas frecuencias incrementáanse enormemente, incluíndo as das terminacións -E (70%) e -N (56%), polo que podemos dicir que son os monosílabos e a súa frecuencia o que lle fornece unha maior variabilidade á correlación DG-PT. Uns exemplos destes monosílabos de alta frecuencia para as terminacións mencionadas son:

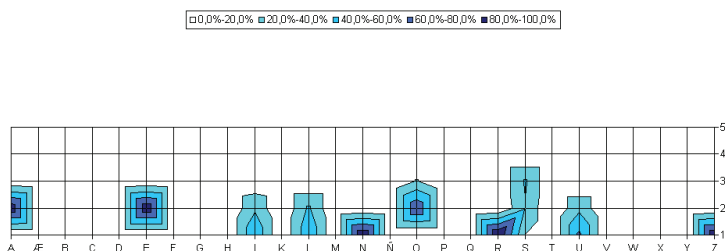
- \* *a, ca, da, dá, lla, ma, pra, xa;*
- \* *e, é, lle, me, pé, que, se, te;*
- \* *ben, can, cen, con, cun, dan, din, don, dun, fan, fin, gran, man, min, nin, non, nun, pon, quen, rin, san, sen, son, tan, ten, un, ven, vin;*

\* *bo, cho, co, do, llo, mo, no, o, ó, pro, só;*

\* *dás, dis, es, fas, grís, has, luns, nos, nós, tes, tras, tres, vas, ves, vés, vos, vós* e os plurais de moitas das anteriores;

Isto confirma a observación inicial e vén corroborado con que a distribución T-SENR-SENM para a terminación -E é a única na que as proparoxítonas aparecen con máis dun 20%. Cando se considera o grupo de estranxeirismos, os cambios máis destacables son, ademais da aparición dalgunhas terminacións predominantemente oxítonas (-K, -W ou -X ou interxeccións en -P), a subida de porcentaxe das paroxítonas para as terminacións -N (59%) e -Y (61%) e a baixada de -I (60%).

Os datos do VOLGA contrastan pola súa maior dispersión (GRÁFICA 12: PT/DG en V-SENR-CONM), pois -I, -L, -O, -S ou -U teñen porcentaxes próximas ou maiores do 20% en máis dunha tonicidade, aínda que este espallamento se traslada ás terminacións -N e -O e -S cando se eliminan os monosílabos.



GRÁFICA 12: PT/DG en V-SENR-CONM

Respecto á presunción de Regueira (2007:243) e a nosa (Fresco 2007:30) sobre que a configuración xeral da posición do acento no galego é de paroxítona para rematadas en vogal ou -S, tódalas distribucións do corpus teatral corroboran a hipótese (aínda que as distribucións sen repetición tamén inclúen o -N), pero os datos do VOLGA xa son discordantes: tenden a ser graves e esdrúxulas as terminadas en -S e a ser agudas as rematadas en -I e -U.

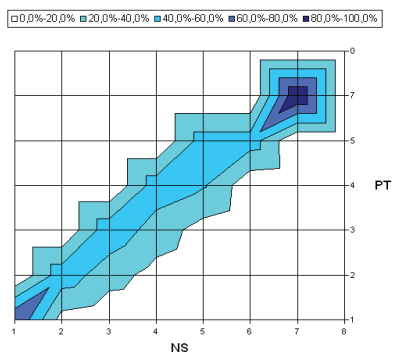
As distribucións de estranxeirismos tenden a asignar coma paroxítonas tamén as rematadas en -Y. Cómpre lembrar que nas terminacións dos estranxeirismos parece haber certa sistematicidade, polo menos cando proceden do francés e tanto se os procesamentos na adaptación procuran “ser fieis na medida do posible á grafía orixinal” como se tentan “reproducir a pronuncia da lingua emprestadora” (Rodríguez Río 1998:306). Isto dá que, sexa que rematen en consoante ou en consoante + vogal muda (onde se elimina ou mantén a consoante final ou cambia a vogal muda), non se modifica a posición da tónica, oxítona no orixinal. Cando os exemplos inclúen linguas, terminacións, modelos, épocas de introdución distintas, os resultados son máis variables:

- a solución para a terminación -ING en inglés amosa varios resultados segundo a época e facilidade (Rodríguez Río 1998:309, nota 24 e 312): *mitin*, *puḍin*, *xilin* vs. *cámping*, *smóking*; aínda así, todos manteñen a vogal tónica como tal (para *footing* e *leasing* resulta máis difícil decidir a posición do acento gráfico, pois a escrita non coincide coa pronuncia);
- aínda que a terminación fonética fricativa palatal sonora pode representarse en galego, pois si se admite en posición implósiva (cf. *ux* do VOLGA e algunhas pronuncias populares de *hoxe* ou *fuixe* vs. Rodríguez Río 1998:306, nota 19), as solucións de *haxix* e *flash* son diferentes pola distinta lingua de orixe;

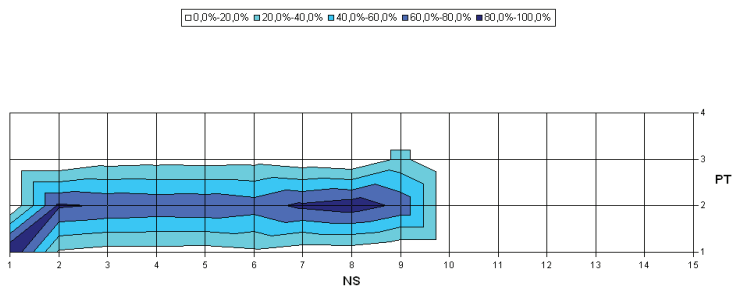
#### 4.6. Distribución NS-PT

O último agrupamento binario relaciona o número de sílabas e a posición da tónica (GRÁFICA 13: NS/PT en T-CONR-CONM). A conclusión obvia á que se chega coa súa comparación é que canto máis se afasta a tónica do final da palabra, máis sílabas ten. Pero disto non se deduce a inversa (GRÁFICA 14: PT/NS en T-CONR-CONM), pois cantas máis sílabas ten unha palabra, non hai máis probabilidades de que a tónica se afaste do final da palabra. Isto implicaría que o galego tería unha certa tendencia a marcar a tónica dende o comezo da palabra (como si pasa noutras linguas coma o

finlandés), pero non é certo. Tamén podemos inferir que cantas máis sílabas ten a palabra, máis probabilidades hai de que sexa paroxítona, contando con que as graves amosan sempre a probabilidade máis alta.



GRÁFICA 13: NS/PT en T-CONR-COMM



GRÁFICA 14: PT/NS en T-CONR-COMM

Estes resultados amosan a particularidade dunha medra de esdrúxulas para as nonasílabas, pero trátase simplemente de que a mostra conta con moi poucos datos (1 esdrúxula dun total de 4 palabras). No VOLGA, a porcentaxe de tónicas nos non estranxeirismos

tamén é máxima para as paroxítonas, pero non é tan acusada. No caso das de 11 sílabas, a porcentaxe de PT repártese a partes iguais entre os dous únicos datos que hai: unha aguda (*radiotelecomunicación*) e unha grave (*esternocleido-mastoideo*).

##### 5. *Desviacións respecto á hipótese inicial e conclusións*

Con esta pescuda puidemos comprobar a veracidade xeral da nosa hipótese de partida e doutras relacionadas. Cos resultados obtidos, queda patente a tendencia do galego a unha alta frecuencia de aparición dos monosílabos, a unha tonicidade en posición maioritariamente paroxítona, ás palabras bisílabas e sobre todo trisílabas na lingua, a unha frecuencia de aparición de fonemas variable en función da posición na palabra, a certa relación entre a categoría gramatical e a forma escrita, a unha dependencia entre a categoría gramatical e o longo da palabra en número de sílabas e a unha correlación entre a tonicidade e os fonemas (e, nunha lingua na que a grafía representa principalmente a fonética, os grafemas) finais.

En concreto, o derradeiro grafema co que se adapta un estranxeirismo si parece influír tamén na sílaba que vai recibir a tónica, aínda que obviamente non a determina. Grazas á análise DG-PT podemos confirmar que en xeral é certa a hipótese de partida (son paroxítonas as rematadas en vogal ou -S e oxítonas tódalas demais), xa que as diferenzas atopadas restrínxense a que non son tan predominantemente paroxítonas (mesmo non minoritarias) as terminacións -E nalgunhas distribucións do corpus teatral e -I, -O, -S e -U noutras do corpus do VOLGA, mentres que a terminacións -N e -L tampouco son tan oxítonas.

Por todo isto, pode propoñerse xa como correcta a aplicación conxunta da hipótese de partida coa metodoloxía de adaptación de arabismos previa. Con todo, cómpre matizar que se debe aplicar antes da eliminación dos grafemas consonánticos (o 1 e o 18 no texto orixinal de Fresco Barbeito 2007:27) no paso da transliteración á transcripción, porque poden crear xuntar vogais que poden ser interpretables coma ditongos ou hiatos na transcripción resultante. Así, por exemplo, a palabra transliterada como *xarii\*a* propónse como

paroxítona antes da perda do grafema <\*> e conservaría a tónica nesa posición, co cal resulta transcrita en galego como *xaría*. De non ser así, cando se aplicase a escolla da tónica ó resultado directo da transcrición, tanto *xaría* como *xaria* cumprirían a posición paroxítona que pediría a terminación en vogal. No caso da simplificación das semivogais porén (Fresco Barbeito 2007:29, punto 27-28) entendemos que se podería aplicar na posición contraria, o cal permitiría reconstruír o orixinal (transliteración *Lìibiya* > transcrición *Libia*, non *Libía*).

A nosa análise pide o contraste de novas hipóteses e abre como sem pre novas portas de investigación: unha análise e proposta sistemática para adaptar ó galego os estranxeirismos que se necesiten en función da lingua de orixe, da cadea fónica e gráfica orixinal e da época de introdución; un estudo da tonicidade e silabación en galego en función de parámetros verbais e silábicos; unha sistematización da adaptación e marcaxe dos estranxeirismos no VOLGA; unha investigación sobre a frecuencia dos fonemas en función da posición na sílaba e o tipo desta; ou unha proposta relativa ó xénero que se lle debe asignar ós estranxeirismos. Esperamos con este pequeno traballo divulgar os nosos resultados e contribuír ó interese da comunidade nestes temas.

### *Agradecementos*

Este artigo nunca sería posible se non fose porque María tivo man de Adrián no tempo no que me dediquei á redacción, Luis e Marisa fixeron o mesmo con nós os tres e Alexandre Rodríguez Guerra tivo a enorme deferencia de mellorar (e moito) a miña iniciativa investigadora.



## Referencias bibliográficas

- Álvarez, Rosario & Xosé Xove (2002): *Gramática da lingua galega*. Vigo: Galaxia.
- Ferreiro, Manuel (1996): *Gramática histórica galega*. Santiago de Compostela: Edicións Laiovento. (Ensaio 64).
- Fresco Barbeito, Víctor (2007) “Transliteración e adaptación do árabe: aplicación terminolóxica”, *Cadernos de Lingua* 29. A Coruña: Real Academia Galega, 5-43.
- Gutiérrez Palma, Nicolás / Alfonso Palma Reyes (2004) “Acento léxico y lectura: un estudio con niños”, en *Revista electrónica de investigación psicoeducativa y psicopedagógica* 4-2(2). Almería: Servicio de Publicaciones de la Universidad de Almería, 143-160. ([http://www.investigacion-psicopedagogica.org/revista/contador\\_Art.php?articulos/4/espagnol/Art\\_4\\_42.pdf](http://www.investigacion-psicopedagogica.org/revista/contador_Art.php?articulos/4/espagnol/Art_4_42.pdf))
- Lakoff, George (1987) *Women, fire and dangerous things. What categories reveal about the mind*. Chicago: University of Chicago Press.
- Méndez, Xosé A. (1997) “Noun Stress in Galician: An OT Perspective”, en Benigno Fernández Salgado (ed.), *Proceedings of the 4th International Conference on Galician Studies = Actas do IV Congreso Internacional de Estudos Galegos: In Memoriam Emilio González López 1903-1991 (Universidade de Oxford, 1994)*, vol. I. Oxford: Centre for Galician Studies, 255-280.
- NOMIG = *Normas ortográficas e morfolóxicas do idioma galego* (2005): Santiago de Compostela: Real Academia Galega / Xunta de Galicia.
- Ohannesian Saboundjian, Maria (2004): *La asignación del acento en castellano*. Universitat Autònoma de Barcelona. Tese de doutoramento inédita. ([http://www.tesisenxarxa.net/TDX/TDX\\_UAB/TESIS/AVAILABLE/TDX-0728105-140941//mos1de1.pdf](http://www.tesisenxarxa.net/TDX/TDX_UAB/TESIS/AVAILABLE/TDX-0728105-140941//mos1de1.pdf))
- Rodríguez Fer, Claudio / Francisco Botana Ferreiro (1984): “Escala da frecuencia fonemática galega” *Grial* XXII/85. Vigo: Editorial Galaxia, 291-302.
- Regueira, Xosé L. (2007) “Patrons fónicos dos nomes de persoa”, en Luz Méndez / Gonzalo Navaza (eds.), *Actas do I Congreso Internacional de Onomástica Galega «Frei Martín Sarmiento» (Santiago de*

*Compostela, 2, 3 e 4 de setembro de 2002*). Santiago de Compostela: Asociación Galega de Onomástica, 239-255 (Biblioteca Galega de Onomástica, 1).

Rodríguez Río, Xusto Alexandre (1998) “O tratamento dos préstamos na norma léxica do galego”, *Verba* 25. Santiago de Compostela: Universidade de Santiago de Compostela, 293-322.

VOLGA = González González, Manuel / Antón Santamarina Fernández (coord.) (2004): *Vocabulario ortográfico da lingua galega*. A Coruña: Real Academia Galega / Santiago de Compostela: Instituto da Lingua Galega. (<http://www.realacademiagalega.org/volga>)