

División de palabras con guiones

Javier Bezos

Compositor y corrector tipográfico

<http://www.tex-tipografia.com>

30 de agosto del 2008

v 4.2

Resumen

La división de palabras con guiones es un problema ortográfico que todavía plantea problemas, tanto desde un punto de teórico como práctico. Es este artículo se analizan los posibles criterios para la división en castellano, con una serie de reglas, y a continuación se estudia la forma en la que se puede implementar un conjunto de tales normas en T_EX. Entre otras se consideran las normas académicas de 1999 y, en especial, del 2005, con comentarios sobre cómo suplir sus deficiencias.

El artículo se divide en dos partes, la primera de las cuales está dedicada al análisis de las reglas. En la segunda se hará una breve historia de los patrones y se tratará sobre su implementación.

1. La teoría

Nuestro primer propósito será delimitar las posibles normas que se puedan aplicar en castellano para la división de palabras. Se trata de un problema que ha sido ampliamente estudiado por numerosos ortógrafos, sin que por por ello se haya llegado a un acuerdo sobre lo que se pueda considerar correcto. Tal vez no sea posible llegar a ningún consenso al respecto debido a que no hay sistema que no presente algún tipo de inconvenientes.

Tales reglas han de formularse de la forma más precisa posible, que prevea los posibles casos y las posibles implicaciones de los patrones generados. Mientras que una división a mano, como se hacía antiguamente, siempre podía ser resuelta

por la intuición del cajista, en la automatización no debe haber lugar ni para la vaguedad ni para la imprecisión, como tampoco para un exceso de reglas basadas en simples consideraciones estéticas (del tipo: «se tiene esta regla, pero si resulta antiestético es mejor aplicar esta otra»). Tampoco hay que excluir posibles casos tan sólo por su infrecuencia y su falta de adaptación perfecta al castellano; las normas han de ser lo bastante precisas para que los términos científicos y técnicos puedan ser tratados adecuadamente. Una docena de reglas puede ser un límite práctico razonable; no parece factible dar menos reglas, teniendo en cuenta que cada una tiene que considerar un caso muy concreto. La selección de las reglas siempre implica cierto grado de arbitrariedad, y, como bien apuntaba Polo [1974], ninguna es completamente satisfactoria.

Las normas

El propósito de este artículo no es proponer nuevas reglas, sino recoger las existentes y darles forma.

La fuente básica debería ser, en principio, la *Ortografía de la lengua española* de la Real Academia Española [1999], intentando dar forma a esos detalles que normalmente han quedado bajo el manto intuitivo de las «sílabas cabales» y rectificando errores evidentes como el que impide dividir *des-truir*. Sin embargo, diversos ortógrafos y ortotipógrafos (particularmente José Martínez de Sousa) ya han mostrado las limitaciones y las deficiencias de las normas expuestas, que son muy vagas, generales y restringidas a casos muy frecuentes.

En el nuevo *Diccionario panhispánico de dudas* [RAE 2005, en adelante DPD], también normativo, se matizan y se extienden muchas de las normas, aunque todavía quedan algunos puntos oscuros pero de importancia menor.¹ Las normas están en el apartado 2 de la entrada *guión* y cada vez que nos refiramos aquí a una sección será de esta entrada. Es interesante destacar que la nueva normativa viene a avalar la línea de trabajo que se estaba siguiendo en los nuevos patrones desde el principio, por lo que se ha tenido un apoyo algo inesperado en aquello en lo que originalmente se desviaron de [RAE 1999].

El principio de que es optativo dividir bien por sílabas, bien por componentes continúa (§ 2.1), pero las reglas que siguen (§§ 2.2-2.6) dan por hecho que la división es por componentes, por lo que la prioridad que en su día se dio a este procedimiento de división en los patrones queda confirmada por las nuevas reglas del DPD. La formulación de las divisiones por componentes en la sección 2.1 sigue de hecho muy de cerca los comentarios que en su día se hicieron aquí: «Esta división

¹Se sigue manteniendo, como en [RAE 1999], que en grupos de tres consonantes, -st- siempre cierra sílaba, lo que una palabra como *destruir* claramente desmiente. En un evidente lapsus, entre los ejemplos de palabras con grupos de tres consonantes está *cons-trucción*.

sólo es posible si los dos componentes del compuesto tienen existencia independiente, o si el prefijo sigue funcionando como tal en la lengua moderna»; como ejemplos de aplicaciones incorrectas de esta norma se dan los tres términos que tradicionalmente se han citado con relación a ella, es decir, *arzobispo*, *península* y *arcángel*.

Según la § 2.2, no se puede dividir nunca entre vocales, salvo si forman parte de dos elementos distintos de una palabra compuesta. Según la § 2.3, no se puede dejar una sola letra a final de renglón.

La sección 2.4 establece que no se puede dividir entre una letra y *h*, salvo si es una palabra compuesta. Los patrones lo permiten actualmente cuando esa letra es consonante (véase más abajo) y así permanecerán, aunque se dará la opción de cambiarlo. En todo caso, afecta a un número muy reducido de palabras, como *adherir*, *inhibirse* o *exhausto*.

Tras la § 2.5, que simplemente dice que la *x* se trata como otra consonante cualquiera, la § 2.6 trata los grupos de consonantes. Aquí hay poco de nuevo, salvo que se recomienda dividir *hiper-realismo*, *inter-racial* y *super-rápido*, como ya estaba en los patrones.

Finalmente, la sección 2.11 añade una serie de recomendaciones, algunas de las cuales también se habían considerado en los nuevos patrones, incluso si no estaban en [RAE 1999], como evitar divisiones que conduzcan a palabras malsonantes y que *-no* quede a principio de línea.

Los criterios

La clasificación de los criterios para la división a fin de línea suele basarse en dos categorías: silábica y etimológica. Uno de los objetivos de este artículo es mostrar las limitaciones de esa clasificación, e intentar crear una nueva que permita matizar otras opciones que esas dos categorías básicas no cubren o lo hacen de forma ambigua. En particular, el principal escollo se encuentra en establecer qué se ha de entender por división «silábica», porque en este caso simplemente estamos pasando el problema a definir qué se entiende por sílaba.

Tal vez el caso más llamativo sea la antigua norma, hoy desaparecida, de que en la división «se ha de formar sílaba cabal» [RAE 1969, pág. 34]. Aquí no sólo hay que intentar interpretar qué significa sílaba, sino dilucidar a qué se refiere con «cabal»: ¿es que existen «sílabas no cabales»? ¿las sílabas son cabales en el contexto de los grupos divididos o de la palabra completa? Obsérvese que *des-es-tim-ar* contiene grupos que, considerados aisladamente, podrían «cabalmente» ser sílabas. Aunque ya se ha abandonado el término, el DPD sigue sin aclarar qué debe entenderse por sílaba. Desde este punto de vista, ni la Ortografía vigente [RAE 1999] ni el DPD dan realmente reglas para la correcta división.

En la actualidad, se considera que la sílaba no tiene existencia fonética real.² La sílaba es un elemento lingüístico que pone de relieve ciertos patrones (fonéticos, métricos, morfológicos y ortográficos) que caracterizan una lengua y en que se basan ciertas reglas como las que veremos más adelante. Si bien es verdad que antiguamente se consideraba que la sílaba influía *per se* en la forma en que se enlazaban los fonemas, con algún tipo de pausa o de variación perceptible, los estudios actuales tienen problemas para acomodar la noción, ya que el habla es de hecho un flujo continuo de emisión de sonidos donde no sólo no hay pausas en el interior de las palabras, sino incluso de una palabra a otra.³ Ciertamente, las eses de *asesor* están tan unidas a las vocales anteriores como a las posteriores.

Por lo expuesto, el sistema silábico debe entenderse como una pura conveniencia basada en reglas a menudo caprichosas que dependen del contexto concreto la lengua y que rara vez pueden justificarse por otra razón que no sea el modo en que tradicionalmente se han mezclado ideas para formar el concepto de unidad silábica. Valga como ejemplo de esto último la palabra «póster» cuya división en español es «pós-ter» y en italiano «po-ster» a pesar de que la diferencia en la pronunciación de las consonantes es mínima y se enlazan con las vocales de idéntico modo.⁴ No es, por sí mismo, un criterio científico.

El criterio etimológico es una especie cajón de sastre donde suelen ir a parar las consideraciones de tipo morfológico, a veces de muy variado tipo, pero que a pesar de todo tienen un punto en común: se basan en la formación de las palabras, tanto desde un punto de vista histórico como de los recursos propios de la lengua para formar nuevas palabras en la actualidad.

Las próximas secciones están destinadas a describir los criterios básicos en los que me basaré: ortográfico, fonético, etimológico, compositivo y derivativo. No cabe duda de que es posible imaginar algún otro criterio, pero aparentemente estos cinco bastan para una descripción completa en la mayoría de las lenguas. Puesto que la división por guiones se trata de un problema ortográfico, consideraré que ese es el criterio fundamental; el resto se basan en consideraciones accesorias —tal vez podríamos hablar de *correcciones*— que lo matizan.

El criterio ortográfico

Con él, establecemos los puntos de división a partir únicamente de las letras que forman la palabra. Como el resto de las normas de ortografía, hay pocas razones

²El Esbozo [RAE 1973], en la página 12, nota 10, no puede ser más explícito: «El concepto de sílaba es más intuitivo que científico».

³A este respecto, es bastante revelador que antiguamente se escribiera el texto sin espacios ni signos de pausa.

⁴En inglés es *post-er*, pero cabe decir lo mismo de la pronunciación.

que sustenten unas posibles normas en preferencia a otras. La ortografía es puro convenio, y todo cuanto podemos aducir es la tradición.

Cuando tratamos con palabras simples nos encontramos con ciertos patrones lingüísticos y fonéticos que tienen su reflejo en la ortografía en la medida que limitan las posibles combinaciones de letras. Así, nunca nos encontraremos con la combinación «*oeu*». El patrón básico que usa el castellano para combinar letras es CV o CVC,⁵ que se pueden encadenar libremente; además, a comienzo de palabra nos podemos encontrar V y VC. Ortográficamente no hay límite en la consonante de cierre, aunque en la práctica la fonética impone que sólo unas pocas pueden tener esa posición.

Este esquema se amplía con la incorporación de las consonantes líquidas y las combinaciones de dos y, en menor grado, de tres vocales, de forma que para abrir podemos tener CC, para cerrar también podemos tener CC, y en lugar de V podemos ver VV e incluso VVV. Las consonantes líquidas (*l, s, r*) sólo pueden aparecer en combinaciones muy concretas vinculadas a la fonética de nuestra lengua. Al tratarse de un problema donde la fonética tiene cierta importancia, aplazamos su tratamiento para más adelante.

Otro aspecto relevante es que la unión de un prefijo y una palabra simple conduce a veces a combinaciones inusuales y atípicas de letras o de sílabas —*morFOFonética, subHumano, deSHacer, nEOEUropeo, BIOElemento*— que sólo se pueden entender en el contexto completo de la palabra compuesta. Por ello, en estos casos resulta imperativo bien conservar la palabra completa o bien dividirla de forma que la particularidad desaparezca, es decir, que se divida por componentes.⁶ Otros casos, como *dirham*, no dejan de ser palabras con grafía extranjera y que no se han acomodado en nuestra ortografía, a veces con pronunciación vacilante (¿la *h* de esta palabra se aspira?).

Un problema accesorio es la simplificación ortográfica de ciertos grupos de consonantes, como en *subrigada*. En principio, no se trata de una simplificación fonética de dos consonantes de igual pronunciación, ya que entonces también debería afectar a *subvención*, por lo que cabe considerar que el objetivo es evitar la presencia del grupo ortográfico *bb* (naturalmente, con independencia de que se pronuncie de una u otra forma).

Podemos establecer nuestras primeras reglas del modo siguiente:

1. Todas las letras se considerarán consonantes menos *a, e, i, o* y *u*.

⁵Por V representaremos las letras *a, e, i, o* y *u*, y por C, el resto. Obsérvese que en este patrón, la *h* sólo aparece en la posición de C.

⁶Sin embargo, Jorge de Buen [2000, pág. 229] sostiene justo el criterio contrario.

2. Salvo en los matices que se introducirán posteriormente, las palabras se dividirán antes de una consonante.
3. Cuando hay varias consonantes reunidas, se dividirá antes de la última de ellas.
4. Nunca se dividirá antes de una o varias consonantes si no hay ninguna vocal después.
5. Nunca se separará una única letra del resto de la palabra.
6. Tampoco se dividirá antes de una *h* ni de la letra que le pudiera anteceder.

La regla 6 ha sido motivo de polémica y probablemente lo seguirá siendo. El principal problema es que con el concepto intuitivo de sílaba tendríamos, por ejemplo, *antia-dherente*; lo tradicional ha sido *antiad-herente*, justificado en parte por razones etimológicas y en parte fonéticas, pero las normas del DPD rechazan explícitamente tal posibilidad. Volveré sobre este asunto más adelante.

El criterio fonético

El principio fonético está basado en el hecho de que una cierta letra puede tener una pronunciación u otra en función de las que le rodean. En estos casos, el criterio fonético persigue mostrar el contexto completo que permita determinar la pronunciación correcta.

Al contrario de lo que ocurre en francés e inglés, no hay ningún caso en español donde una vocal modifique su valor fonético según la consonante que le siga. Sí puede ocurrir que se modifique un sonido consonántico; por ejemplo, en *envío*, la *n* tiene el sonido de la *m* por ir seguida de *v*, mientras que en *encontrar*, la *n* se velariza. Este ejemplo muestra que la aplicación estricta del criterio fonético puede presentar algunos problemas y por tanto resulta necesario cierto grado de flexibilidad: no parece prudente impedir *en-vío* o *en-contrar*, debido a que el cambio fonético no es realmente substancial, es decir, es simplemente una variante fonémica y cualquier hispanohablante entenderá la palabra correctamente se pronuncie como *n* o como *m*.⁷

Una excepción obvia, e importante, es el grupo de letras que representan un único fonema, como *ch* y *ll*. No son los únicos ejemplos, porque en *washingtoniano* el grupo *sh* también puede tener valor fonético propio.

7. Los grupos consonánticos que representan un único sonido, principalmente *ll* y *ch*, son indivisibles y se tratarán como si fueran una única consonante.

⁷En la formación de compuestos puede haber caída de consonante, pero tiene reflejo gráfico: de *in- + legal* obtenemos *ilegal*.

Otro caso especial es *rr*, debido a que la *r* sola puede representar el mismo sonido, que además es básicamente una prolongación del sonido simple (parecido a *-nn-*); también puede tener una prolongación adicional en compuestos como *inter-regional*, «ya que, en estos casos, la graffa *rr* no representa un solo sonido, sino dos, uno vibrante simple y otro vibrante múltiple» (DPD, pág. 553). Por tanto, no se trata del mismo caso que los grupos de la regla anterior. Aceptaremos la tradición de conservar unido este grupo, aunque no de forma tajante. Volveré sobre ello al tratar la división compositiva.

Otra excepción importante es la letra *y* que es consonante o vocal según el contexto: *reyes* frente a *rey*. Como la *y* es consonante cuando no va seguida de consonante y vocal en el resto de los casos, no sería necesario añadir ninguna regla adicional; un análisis cuidadoso revela que bastaría considerar que es consonante (como se ha hecho en la regla 1) para obtener los resultados correctos, si no fuera por un pequeño detalle: la regla 4 impediría una división como *Salabe-rry*.

8. La *y* se considera consonante cuando no va seguida de consonante, y vocal en el resto de los casos.

Esta regla tiene una pequeña trampa, ya que en las terminaciones *ay*, *ey*, etc. la *y* no tiene valor de consonante. Sin embargo, no tiene relevancia a efectos prácticos. Nótese también que en esta discusión he obviado el hecho de que la *y* puede ser semiconsonante (o semivocal).

El español tiene una característica fonética singular con relación a otras muchas lenguas: la diferenciación entre diptongos e hiatos. Aquí tenemos un ejemplo de cómo una cierta peculiaridad fonética tiene su reflejo en la silabificación. Naturalmente, si dos vocales se articulan separadas no es porque haya una frontera silábica entre ellas; más bien al contrario, la abstracción que supone el concepto de sílaba impone una frontera precisamente allí donde hay un hiato. Para saber de antemano si una vocal va unida a la siguiente *conviene* tener a la vista todo el grupo vocálico, por lo que introducimos una nueva regla:

9. No se dividirá entre vocales.

Subrayo la palabra *conviene* porque no se trata de un cambio sustancial del sonido de las vocales, y por tanto no debe considerarse una necesidad imperativa que se impida la división entre ellas, como lo demuestra el hecho de que puede darse un diptongo incluso entre palabras. Por otra parte, es preciso recordar que no hay reglas ortográficas que permitan distinguir un hiato de un diptongo, lo que hace que sea todavía más importante tener todas las vocales sin separar.⁸

⁸Las reglas que se suelen dar al respecto pretenden tan sólo dar un criterio sencillo para la acen-

En cambio, sí es frecuente que una consonante tenga un sonido que depende de la vocal que le sigue: *cal/ci*, *ga/gi*. No hay que considerar ningún caso especial aquí, pues en la división ortográfica la vocal está unida siempre a la consonante que le antecede, y no hay prefijos que puedan dar a grupos así en la frontera.

Si bien, como he dicho, el silabeo de las palabras no tiene existencia fonética propia, sí influye en ciertas características que se pueden entender tan sólo desde la lógica de la lengua considerada. Por ejemplo, en castellano la *d* va unida regularmente a la siguiente vocal, pero cuando tratamos con formaciones etimológicas en palabras como *adherir*, donde la *h* es muda, puede producirse el fenómeno de considerar la *d* como coda silábica, lo que ciertas regiones implica un cambio al sonido de la *z*.⁹ Un fenómeno paralelo, y sin duda mucho más extendido, es la permanencia del valor semiconsonántico de la *u* en *deshuesar* (cf. *desuello*).

Ciertas consonantes, llamadas líquidas, pueden unirse a otras de forma que se articulan de una sola vez. Este suele ser el caso de *bl*, *br*, *cl*, *cr*, *dr*, *fl*, *fr*, *gl*, *gr*, *kl*, *kr*, *pl*, *pr*, *tr*; pero hay que destacar que no toda combinación de esas consonantes se articula unida, de forma parecida como ocurre con las vocales: en particular, los grupos *bl*, *br* y *dr* pueden separarse.¹⁰

10. Las letras *l* y *r* permanecerán unidas a la consonante que le antecede cuando son líquidas, y la posible división será antes del grupo formado.

Esta regla, al imponer el valor fonético líquido de *l* y *r*, excluye palabras como *postludio*, donde la *t* normalmente es muda y que se dividirá aplicando las reglas generales vistas antes: *post-ludio*. En realidad, este grupo presenta el problema fonético de que en España su *l* no es líquida mientras que en América sí lo es; por tanto, se evitará la división (aunque en un futuro pueda dejarse como opcional).

Al componer palabras pueden aparecer otras anomalías prosódicas. Algunos de los ejemplos vistos antes (como *neoeuropeo* y *bioelemento*) pueden valer también aquí, pero están siempre relacionados con la composición de palabras.

Con todo, hay casos que quedan por completo fuera del análisis anterior, ya que hay circunstancias excepcionales que determinan de una forma insospechada la pronunciación. Si al final de una línea nos encontramos con *subrepti-*, estamos todavía sin la información necesaria para la pronunciación del grupo *br*; sólo la encontraremos cuando veamos la siguiente línea, es decir, cuando veamos lo que hay

tuación de las palabras y en modo alguno se pueden considerar un reflejo de la articulación real de los grupos vocálicos. Así, *piano* suele tener hiato en *ia*, y *arqueólogo*, diptongo en *eo*

⁹[Bosque y Demonte 1999, vol. 3, pág. 4452]. Los fonetistas lo suelen considerar una falta de prosodia, pero mi objetivo aquí no es tomar partido por el uso o por el estilo, sino tan sólo constatar la realidad.

¹⁰Según el DPD, § 2.6b, es así porque «cada consonante pertenece a una sílaba distinta», aunque más bien debería decir que pertenecen a sílabas distintas porque se pronuncian separadas.

¡cinco letras más adelante! (*su-brepticio* frente a *sub-reptiliano*, por ejemplo).¹¹ O bien dejamos la palabra completa para que el lector tenga todos los datos, o bien dividimos de forma que se ponga de relieve la forma correcta (tal y como hemos hecho hace un momento).

Los criterios etimológico, compositivo y derivativo

Debido a la estrecha vinculación de estos tres casos, los trataré en un único apartado.

El criterio etimológico histórico es aquel en el que los recursos por los que la palabra se ha formado ya han desaparecido y no son productivos en la actualidad, incluso si el hablante puede tener conciencia de ello: prefijos o raíces que han caído en desuso, variación de la forma de la palabra con el tiempo, etc. Por ejemplo, *adherir*, *arz-obispo*, *pen-ínsula*, *con-struir*, serían divisiones etimológicas.

El criterio derivativo separa los elementos que sirven para modificar la estructura de las palabras. La división derivativa es frecuente en inglés (*auto-mat-ic-al-ly*), aunque en el inglés americano se mezcla con la fonética (*auto-mat-i-cal-ly*).

El criterio compositivo se aplicaría a las voces formadas a partir de otras simples. No está claro si en español la prefijación se puede considerar derivación o composición, ya que participa de características de ambos. Sin embargo, en el uso de estos términos no he intentado establecer una correspondencia rigurosa con los conceptos morfológicos.

Los siguientes ejemplos pueden ser esclarecedores: *per-spicaz* sería una división etimológica ya que no es una palabra compuesta con elementos claramente individualizados, *flor-ec-illa* serían divisiones derivativas, pues la palabra ha sido formada por el hablante a partir de sufijos e infijos conocidos, mientras que *sub-reino* es una división compositiva.

Debido a que la división etimológica no aporta en castellano ningún dato realmente relevante y puede alterar la prosodia de forma significativa, no ha tenido ningún uso y por tanto no la consideraré. Con todo, algunas palabras conservan ciertos grupos etimológicos (*dismnesia*, *metempsychosis*), y nos encontramos con un caso, tal vez el único, donde la división etimológica es insoslayable. Estos grupos suelen estar asociados a palabras compuestas, pero como también se pueden dar en palabras simples o lexicalizadas añadimos la siguiente norma:

11. Los grupos *pt*, *ct*, *cn*, *ps*, *mn*, *gn*, *ft*, *pn*, *cz*, *tz*, *ts* precedidos de consonante, se dividirán entre esa consonante y el grupo.

¹¹ Algunos autores consideran incorrecta la pronunciación que doy aquí de *subrepticio*, pero no pretendo tomar partido por el uso o por el estilo.

Esta regla se aparta del criterio silábico tradicional menos de lo que parece. Consideremos la palabra *fenolf-taleína*, donde aparece la combinación insólita *lft*; aunque en principio la regla 3 apunta a *fenolf-taleína*, en realidad una división por sílabas debería dar prioridad a *ft* como cabeza silábica, que encontramos en *ftalato*, frente a *lf* como coda silábica, que es ciertamente raro.¹² Esto explica también por qué he impuesto la condición adicional de la consonante previa al grupo: mientras que parece que estamos más acostumbrados a usar como coda silábica cualquier letra (sobre todo cuando es final: *currículum*, *complot*, *reloj*) parece más difícil aceptar combinaciones como *sm* o *mp*. Naturalmente, pueden darse combinaciones aceptables, pero en un caso tan específico y escaso como el contemplado por esta regla, parece conveniente adoptar una formulación simple y directa, sin «subcasos» especiales (que se tratarían como excepciones, como en *tsetsé*, *plancton* y *sancta sanc-tórum*); esta regla también resuelve algún derivado como *hertziano*. Sin embargo, recuérdese que en palabras compuestas siempre es posible la división compositiva, como en *para-psicólogo*.

La derivativa puede aportar más datos, pero la presencia de interfijos (como *-ec-*) y el hecho de que la flexión no se basa en la palabra completa pueden introducir una complicación adicional que no queda compensada por las posibles ventajas (si las hubiera). Por contra, la división compositiva tiene tradición, no es difícil de aplicar y coincide, como ya se ha visto, con anomalías de prosodia, sin que por ello aparezcan nuevas irregularidades.¹³

En la división compositiva es necesario establecer una distinción entre palabras compuestas en sentido estricto y palabras con prefijos. Ejemplos de las primeras son *hispanoamericano*, *ciudadrealeno* y *cientificotécnico*, pero no estoy muy seguro de que la primera de estas palabras entre en esta categoría (es decir, que *hispano-* no sea un prefijo) y grafías como la última son raras y se prefiere intercalar un guión. Obsérvese que el ejemplo visto antes de *post-ludio* tiene todo el aspecto de una división compositiva, gracias al prefijo *post*, pero *ludio* revela que en realidad es una formación etimológica.

Con relación a la frontera entre prefijo y raíz, hay que considerar cuatro casos, según se trate de vocales o consonantes:

- *V-C* y *V-CC*. Aunque parezca que es el caso más simple, porque el criterio ortográfico parece encajar perfectamente, de hecho nos encontramos con al-

¹²El hecho de que muchos autores den *fenolf-taleína* como la división silábica correcta, revela hasta qué punto es vago el concepto intuitivo aplicado y cómo la «norma» tiene en realidad muy poco de norma.

¹³Las lenguas finougrias pueden considerarse una especie de combinación de compositiva y derivativa, ya que los sufijos pueden tener idéntica función que nuestras preposiciones. Como se ve, la elección de un sistema u otro depende de la naturaleza de la lengua.

gunas complicaciones.

La primera de ellas es cuando la raíz comienza con *r*, en cuyo caso se duplica ortográficamente. Aquí podemos preguntarnos si los elementos que inicialmente estaban divididos, se unen y se vuelven a dividir, o si directamente permanecen sin unir porque el prefijo ha quedado en una línea y la raíz en otra. En el primer caso se deberían respetar los posibles cambios; en el segundo, se preservaría la grafía original de los elementos de forma que tenemos *extra-radio*. La siguiente pregunta sería si hay alguna ventaja intrínseca en un punto de vista u otro, pero no creo que lo haya, con la excepción de que hubiera que tratar a mano palabras afectadas.

Cuando la raíz comienza con *h* debemos tener en cuenta que estará entre vocales, y por tanto lo dicho en el apartado V-V se puede extender a este caso.

El grupo inicial CC suele tener la segunda consonante líquida, pero hay ciertos grupos que aparecen en palabras técnicas: *pterodáctilo*, *ctenóforo*, *cneoráceo*, *psique*, *mnenónico*, *gnato*, *flalato*. Un problema adicional es si estos grupos deben respetarse incluso si no van precedido de un prefijo productivo o se trata de una palabra lexicalizada (*ortog-nato*); la respuesta a este problema no tan complicada como parece, ya que el principio ortográfico sigue siendo válido; sólo cuando uno de esos grupos está precedido de consonante debemos establecer reglas adicionales, tal y como ya hemos visto.¹⁴

- *C-V* y *CC-V* Esta división, aparentemente *contra natura*, es menos conflictiva de lo que parece. Ya hemos visto que la pronunciación de una vocal nunca queda alterada por las consonantes adyacentes y viceversa, y por ello no presenta problemas especiales.
- *V-V*. Dos vocales átonas siempre forman diptongo, lo que podría interpretarse *a priori* como un obstáculo para la división compositiva. Sin embargo, la diptongación también es posible entre palabras sin que por ello se escriban juntas; por tanto, la separación se puede considerar igualmente válida, incluso dentro de palabras, cuando se trate de una frontera morfológica clara.¹⁵ En palabras lexicalizadas se han simplificado las vocales dobles (*rescribir*), lo que queda fuera de la división compositiva y entraría dentro de la etimológica, de forma que el único criterio aceptable es el ortográfico.

¹⁴Una persona acostumbrada a palabras con el grupo *gn-* me ha comentado que ve poco natural separarlo, por lo que me pregunto si esta regla, o en general cualquier otra, no podría depender de si el escrito es técnico o no.

¹⁵Podemos comparar *anti-Alemania* con *antialemán*, como se verá más adelante.

- *C-C*, *CC-C* y *C-CC*. Los prefijos que terminan en consonante son pocos, pero entre ellos están los más frecuentes: *in-*, *des-*, *super-*... Tanto la *n* como la *s* son fonéticamente invariables, y se pueden unir tanto a la vocal que le antecede como a la que le sigue. No ocurre lo mismo con la *r*, ya que al añadir un prefijo siempre se conserva su sonido doble propio de la posición inicial sin que tenga reflejo gráfico. Las reglas ortográficas ya consideran el caso cuando el prefijo termina en *n* o *s*, pero si el prefijo acaba en otra consonante, como *sub-*, *post-* e *inter-*, no se refleja ortográficamente la pronunciación real.¹⁶ El caso *CC-C* puede presentarse en *postromántico* si aceptamos tal grafía en lugar de *posromántico*; puesto que el objeto de los patrones no es condicionar la ortografía de las palabras, tales combinaciones deben ser consideradas. Lo mismo cabe decir de *trans-oceánico*. Por tanto, en los casos de la *r* y de *sub-* es particularmente importante ayudar al lector dejándole clara la presencia del prefijo, si no se tiene toda la palabra a la vista.

De esta forma, tenemos nuestras últimas reglas:

12. Las palabras compuestas de prefijo y raíz, cuyo significado corresponda a los elementos separados, y siempre que en su formación no se alteren sus formas, se dividirán por sus componentes.
13. El grupo *rr*, cuando es parte de una palabra simple, es indivisible y se tratará como si fuera una única consonante.
14. A cada componente de las palabras compuestas se le aplicarán por separado las reglas generales de división de palabras.

La regla 15 implica que *antie-conómico* no es válida, al aplicar la regla 5 a la palabra *económico*. En cuanto a la 14, la *rr* se tratará por la regla 13 cuando sea producto de la unión de dos *r* al componer una voz.

Un problema adicional, de difícil solución, es la formación de prefijos a partir de verbos: *corta-*, *guarda-*, *limpia-*, *cubre-*.

Acabo este apartado con unas pocas reglas accesorias. La primera, muy curiosa y que se ha venido observando regularmente, es impedir palabras *malsonantes* en la división (*com-puta*, *tubér-culo*, *cuadrú-pedo*). A decir verdad, muchas de ellas responden a pura mojigatería y no sé si es oportuno considerarlas (*culo*, *pedo*), pero tal vez sí haya que tener en cuenta algunas otras, cuando el resultado sea directamente la palabra que pueda ofender al lector (*com-puta*, pero no *comp-putar*). Otra regla impide dividir palabras de cuatro letras, como *pe-na*, *li-no* y *co-sa*. Finalmente, si aparece a principio de línea una sílaba *no* final, no es raro

¹⁶Lo que origina errores en los hablantes, ya que no es raro oír *subrayar* con *r* líquida y no, como sería lo correcto, doble.

que se *relea* y por tanto se cambie el sentido de la frase («el Gobier- | no aprobó el decreto»).

Comentarios sobre algunas fuentes

Pocos autores han estudiado tan a fondo el problema de la división de palabras como José Martínez de Sousa. Su *Diccionario de ortografía de la lengua española* [1996] es una fuente esencial para cualquiera que quiera profundizar en el tema, tanto por la multitud de fuentes citadas como por la indudable experiencia de alguien que conoce los aspectos teóricos de la ortografía y los prácticos de la tipografía. Sin embargo, rechaza ciertas normas tradicionales porque no se adaptan a cierta lógica, cuando ya he dicho que la ortografía es ante todo arbitraria. Veamos algunos puntos.

[...] si se divide *sub-/oficial* se atenta contra la prosodia, pues lo que se pronuncia es *su-bo-ficial* y no *sub-oficial* (pág. 174); [...]

La opinión de este autor se basa probablemente en la idea de que el silabeo modifica esencialmente la forma en que se enlazan los fonemas y que el guión señala algún tipo de separación perceptible. Sin embargo, tal planteamiento presenta algunos problemas: si dijéramos que «Hitler quería crear una super-Alemania con superalemanes» resulta patente que la presencia del guión no altera la forma en que se enlazan la *r* y la *a* y que obedece únicamente a la necesidad ortográfica de la mayúscula en el nombre propio. Por tanto, *su-bo-ficial* y *sub-oficial* tampoco tienen diferencias de pronunciación, y no la tendrían incluso si los elementos se escribieran por separado (en lo que respecta a la unión de fonemas). Un par de ejemplos pueden aclarar este punto: *mal educado* y *maleducado* —donde vemos diferencias en la acentuación y la entonación, pero no en el enlace *le*—, y el vulgarismo *un amoto, del amoto*.¹⁷

Ramos Martínez en su *Corrección de pruebas tipográficas* se centra en el problema de las voces compuestas e intenta dar sentido a las normas ortográfica de la Academia del año 1959. Por ejemplo, con relación a *arzobispo* dice:

Si de *archi* sale *arqui* y de éste *arz*, son perfectamente analizables. (pág. 93)

Aunque *obispo* sea claramente reconocible, *arz* ya no es activo (a pesar de que aún figura en el Diccionario académico) y por tanto esta división ha de considerarse etimológica. De hecho, este autor mezcla muy libremente las divisiones etimológicas, las compositivas, e incluso las derivativas:

¹⁷Si efectivamente existieran esas marcas delimitadoras en sílabas o palabras, este vulgarismo carecería de explicación, porque la *a* se separa de la consonante que le precede para pasar ¡a otra palabra! Como no existen, la *o* final provoca ese desplazamiento.

Tal es el caso del sufijo *itis*, que significa *inflamación*, cuya división debería ser: *neur-itis, celul-itis*; [...] (pág. 97)

También rechaza la división *no-so-tros* porque «*so* no es analizable aunque sea pronunciable»; de ser coherentes, eso implicaría que la única división válida es la etimológica. Puesto que la norma ortográfica especifica claramente que su aplicación se restringe a palabras derivadas con prefijo más raíz, es indudable que este autor extiende la norma más allá de lo que debería.

Julio Casares fue uno de los artífices de las normas ortográficas de los años 50, y en el «Informe» de las normas de 1952 dice:

¿Es juicioso imponer al que escribe estas y otras muchas preocupaciones por el accidente puramente mecánico de que se le haya acabado el espacio antes de terminar una palabra? (pág. 94)

Aquí coincidiría con Casares si no fuera por el sentido que en realidad quiere darle, porque en la página siguiente añade que se busca

una regla que sea de fácil observancia para todo el que hable español y cuya aplicación no presuponga conocimientos especiales.

Claramente, Casares quiere dar solución al problema práctico que se presentaba a los cajistas en una época en la que la división se hacía a mano. Pero actualmente hay que ir pensando más en la conveniencia del lector y no del compositor, pues es posible automatizar casi por completo esta tarea enojosa sin que el compositor o maquetista tenga que poseer conocimientos especiales: hoy se ha realizado el deseo de dividir palabras de forma puramente mecánica sin que por ello se comprometa la coherencia, la tradición o la legibilidad.

A pesar de ello, en el DPD la Academia acaba de prohibir divisiones como *adherir*; se ha venido haciendo así desde hace décadas y suprimir esta posibilidad es puro escolasticismo sin soporte en ninguno de los criterios vistos (salvo, tal vez, el «estético»). De hecho, las normas anteriores a 1952 no incluían tal posibilidad y en éstas se añadió a sugerencia de Casares por ser «fácil de practicar por el más lego» [RAE 1952, pág. 96]. Aquí, la intención de Casares tiene pleno sentido.

El análisis de José Polo en *Ortografía y ciencia del lenguaje* es uno de los más extensos y detallados, y al tiempo uno de los más deficientes. Es quien más se toma al pie de la letra el nombre de división etimológica y por tanto quien más la confunde con la compositiva. Además, claramente desconoce lo que implica la tipografía y desprecia conceptos básicos de legibilidad:

A veces pensamos que, en el ámbito de lo impreso, puede dársele una importancia excesiva a cuestiones secundarias, como más o menos espacios, tipo algo menor, etc., cosas que el lector normal ni percibe. (pág. 173)

Obviamente, el lector las percibe aunque no sea consciente de ello.

Polo comienza dando preferencia a la división silábica por lo que él considera complejidades de la etimológica, pero admite que no se puede aplicar directamente y comienza a añadir excepciones. Finalmente, puesto que ninguno de los dos métodos le satisface, termina con la radical propuesta de no dividir palabras nunca; es evidente que no sabe cómo dar una solución al problema y por ello opta por la vía rápida de ignorarlo:

¿Compensa tanta excepción o matización de lo que, en principio, parecía una norma sin suturas? Nuestra opinión es decididamente favorable a no separar las palabras al final del renglón [...]. (pág. 173)

Polo demuestra ser un agudo observador pero un analista muy superficial, lo que lastra esta parte de su obra (y de hecho, toda ella).

2. Los patrones de T_EX

Los problemas relacionados con la generación de patrones de T_EX ya han sido ampliamente estudiados por Petr Sojka, que ha publicado varios artículos donde se discuten las limitaciones del sistema de T_EX. Es curioso que mientras que en otros otros sistemas se considera que el algoritmo de T_EX es bastante avanzado, en el mundo de T_EX ya se considera muy limitado.

En nuestro caso, se sigue un sistema en dos pasos para permitir crear archivos de patrones en diversas codificaciones. Tendríamos así un archivo escrito con algún tipo de marcas que nos permitan decidir la forma final del patrón, del que obtendríamos un archivo de T_EX adaptado a nuestras necesidades.

Quiero destacar la amabilidad de Francesc Carmona y su disposición para colaborar en este proyecto para unificar patrones. Su trabajo se ha tomado en consideración y de él se han tomado directamente algunos patrones.

Otros patrones en el directorio `spanish` de CTAN se eliminaron, ante la imposibilidad de contactar con los autores, quienes en algunos casos, como en el de Julio Sánchez, nos constaba que ya no estaban implicados con el mundo de T_EX. Además de Carmona, Sánchez y García (mencionado más abajo), teníamos los de Aurion Tecnología (México) y otro más anónimo, facilitado por Mario Widel. Recientemente, se han dejado en CTAN sólo los nuevos, pero cuando parecía que se aclaraba definitivamente la situación, han aparecido otros nuevos patrones, de Rodrigo Medina, que no aportan nada nuevo y adolecen de los mismos problemas que los que se acabaron eliminando.

El generador de patrones

La elección del lenguaje para el generador de patrones no ha sido difícil. Julio Sánchez optó por shell y Ramón García por C, pero yo he preferido usar directamente T_EX, que después de todo es un lenguaje de programación relativamente completo que permite escribir en archivos. También he considerado otras opciones como JavaScript (que deseché porque no podía crear archivos) o Python.

Al contrario que los patrones en sí, el archivo con la base usada para su creación es puro texto ASCII con caracteres especiales para precisar el comportamiento del patrón. De esta forma, los prefijos vienen indicados con / . . . / y las letras acentuadas van precedidas de ' , salvo en los patrones que deben ir duplicados con y sin acento, donde se usa ". Este último caso se da en palabras como *desempañar* (des3emp"a'n), donde puede aparecer algún acento si se añade enclíticos (*desempañalo*) [Éste último no está implementado todavía, sino que están añadidos a mano].

Las palabras consideradas son sobre todo las incluídas en el diccionario *ispell* de Santiago Rodríguez y Jesús Carretero, basado a su vez en un *corpus* de 20 millones de palabras compilado por los autores. El léxico básico del diccionario contiene más de 80 000 voces distintas, con sus correspondientes reglas de derivación. Además, he podido crear una colección de casi 700 000 palabras, que incluyen un buen número de verbos completamente conjugados y palabras derivadas.

Los patrones silábico-ortográficos se forman con los códigos 1 y 2, y sólo tienen consonantes, por lo que con ellos tendríamos una división ortográfica con cualquier codificación.¹⁸ Otros patrones prefieren basarse en combinaciones de consonante + vocal, pero en tal caso estaríamos restringiendo sin necesidad los patrones a una cierta codificación de fuentes (que normalmente es T1). La regla 3 es la que nos permite conseguir fácilmente la división silábica sólo con consonantes: se autoriza dividir antes de cualquier consonante (por ejemplo, 1b), y luego se impide antes de dos consonantes (2bc) salvo con las líquidas (b2r, a la que hay que añadir 2br. para evitar sorpresas a fin de palabra).

Los patrones etimológicos se tratan igual que los ortográficos, pero ya tenemos que pasar a códigos 2 y 3. Sólo se consideran algunas combinaciones probables (por ejemplo, no se incluye *qpt*).

Los patrones de prefijos también se basan los códigos 1 y 2, pero las excepciones pasan a 3 y valores superiores. En muchos casos se podría haber reducido el 4 a 2, pero he preferido mantener conjuntos de valores separados para mayor claridad. La entrada de prefijos se ha automatizado, de forma que con escribir /prefijo/ ya se crean todos los patrones básicos.

¹⁸Recuérdese que nunca se divide entre dos letras a menos que un patrón lo autorice. Un error recurrente ha sido añadir patrones innecesarios para evitar divisiones.

Para evitar cierto número de divisiones no deseadas en verbos como *contraer* o *superar* se han añadido las terminaciones verbales correspondientes, incluyendo las del voseo de buena parte de América (Argentina, Paraguay, Uruguay, Perú, Venezuela, Nicaragua, El Salvador, etc.). Para ello, se ha consultado el interesante cuadro del DPD págs. 675-676.

El archivo generado

Como ya he dicho, el objetivo es permitir cierta flexibilidad en el archivo generado. Sin embargo, en el primer paso sólo se ofrece la salida en T1. Obviamente, eso implica añadir el código necesario para que T_EX pueda leerlo correctamente. No recorro a ningún paquete sino que tan sólo añado las macros mínimas para que se pueda leer el archivo

Algunos problemas

Aunque la regla 13 es deseable y por tanto se ha tenido en cuenta, ha sido necesario crear patrones en los que cada prefijo aparece combinado con una las 10 vocales (con y sin acento), así como ciertas consonantes en casos concretos:

- La *h* en todos los casos. [[[¿Hace falta?]]]
- La *r* si el prefijo acaba en *b* o en *r*.
- la *l* si el prefijo acaba en *b*.

El problema es que cada prefijo necesita 11 patrones como mínimo, y de medio centenar se crean más de medio millar de patrones, con el evidente gasto de memoria.

Los prefijos breves son problemáticos. Podemos estar seguros de que *aero-* puede quedar bajo nuestro control, pero intentar establecer patrones para *des-* resulta más complicado: es indudable que la combinación de esas tres letras aparece con tanta frecuencia que difícilmente vamos a poder considerar todos los casos. Por ello, me parece preferible evitar en general la división, y que si divide sea en unos pocos casos en los que podemos estar seguros de que se dividirá bien. Los prefijos así tratados son: *a*, *an*, *co*, *in*, *des*, *re*, *en*, *sub*, *pre*.... No debe sorprendernos que *desalar* sea indivisible, debido a que *de-salar* y *des-alar* han quedado excluidas, y a que *desa-lar* violaría la regla 5 (por medio de la 13) si la formación fuera *des-alar*.

Para el caso de *des*, la guía principal ha sido ver si existen las ternas *des/re/()* (semánticamente paralelas): *desoír/reoír/oír*, *desalar/resalar/salar*, pero no *desarrollar/rearrollar/arrollar* y *desalar/realar/alar*.

Patrones adicionales

No tendría sentido incluir en la lista básica de patrones una serie de prefijos de escaso uso en condiciones normales; se trata de términos de uso técnico en diferentes ramas, y por ello se separan del archivo principal para que se cargen opcionalmente. De momento no se incluirá más que un tanteo para química.

3. Bibliografía

- Ignacio Bosque y Violeta Demonte, *Gramática descriptiva de la lengua española*, Madrid, Espasa-Calpe, 1999.
- Jorge de Buen, *Manual de diseño editorial*, México, Santillana, 2000.
- José Fernández Castillo, *Normas para correctores y compositores tipógrafos*, Madrid, Espasa-Calpe, 1959.
- Real Academia Española y Asociación de Academias de la Lengua Española, *Diccionario panhispánico de dudas [DPD]*, Madrid, Santillana, 2005.
- Real Academia Española [RAE], *Nuevas normas de prosodia y ortografía*, Madrid, Imprenta Aguirre, 1952.
- Real Academia Española [RAE], *Ortografía*, Madrid, Imprenta Aguirre, 1969.
- Real Academia Española [RAE], *Ortografía de la lengua española*, Madrid, Espasa, 1999.
- Real Academia Española [RAE], *Esbozo de una nueva gramática de la lengua española*, Madrid, Espasa-Calpe, 1973.
- V. Martínez Sicluna, *Teoría y práctica de la tipografía*, Barcelona, Gustavo Gili, 1945.
- José Martínez de Sousa, *Diccionario de ortografía de la lengua española*, Madrid, Paraninfo, 1996.
- Juan José Morato, *Guía práctica del compositor tipográfico*, Madrid, Hernando, 2.^a ed., 1908 (1.^a ed., 1900, 3.^a ed., 1933).
- José Polo, *Ortografía y ciencia del lenguaje*, Madrid, Paraninfo, 1974.
- R. Ramos Martínez, *Corrección de pruebas tipográficas*, México, UTEHA, 1963.
- P. Sanmartí, *Compendio de gramática castellana*, [Barcelona, s/e, 1907].