



Vietlex have developed this website for freely access to all people who want to study, research, and understand linguistics generally and Vietnamese specifically. Vietnamese Corpus of Vietlex with more than 150.000.000 syllables.

All home and foreign collaborations and cooperations with Our Centre are welcome !

[Giới thiệu](#) [Trang chủ](#) [Ngôn ngữ học](#) [Xử lý ngôn ngữ](#) [Sản phẩm](#) [Đặt mua sách](#) [Kho ngữ liệu](#) [Tra từ điển](#) [Lịch vạn niên](#)

## Xử lý ngôn ngữ

## Bài trong chuyên mục

### Quy tắc đặt dấu thanh trong tiếng Việt

VŨ XUÂN LƯƠNG - Trung Tâm Từ Điển Học

Tiếng Việt được thể hiện qua hai dạng chữ viết, chữ Nôm và chữ quốc ngữ. Các vấn đề nghiên cứu để biểu diễn chữ Nôm trong công nghệ thông tin mới được nghiên cứu và gần đây đã thu được những thành công bước đầu. Một quan tâm chính của giới công nghệ thông tin trong những năm qua là tập trung nghiên cứu chữ quốc ngữ để biểu diễn bên trong máy tính. Kết quả của quá trình nghiên cứu là đã công bố được một bảng mã tiêu chuẩn cho tiếng Việt, viết tắt là TCVN-5712 (1993). Cùng với bảng mã TCVN-5712 là sự ra đời của bộ font chữ ABC, được quy định là tiêu chuẩn quốc gia. Tuy nhiên, do còn nhiều hạn chế mà bảng mã TCVN-5712 không được sự ủng hộ của giới làm phần mềm, và kết quả là trên lãnh thổ Việt Nam cũng như ở nước ngoài đã sản sinh ra rất nhiều bộ mã cho chữ Việt, dẫn đến một tình trạng hỗn loạn chữ Việt trên thị trường công nghệ thông tin. Việc ra đời của nhiều phần mềm gõ chữ Việt có tích hợp nhiều cách bỏ dấu khác nhau cũng góp phần tạo nên sự hỗn loạn đó. Hậu quả là gây trở ngại không nhỏ cho việc lưu trữ, xử lý cũng như trao đổi ngữ liệu trên máy tính cá nhân, trên mạng máy tính, trên Internet, v.v.

Để tránh sự lệ thuộc vào những thói quen mang tính chất cá nhân, thiết nghĩ việc xác định vị trí đặt dấu thanh cho chữ Việt cũng cần được chuẩn hoá, làm cơ sở cho những người các phần mềm gõ chữ Việt theo đó mà có giải pháp xử lý phù hợp.

#### Định nghĩa

1. **Âm tiết** là đơn vị phát âm tự nhiên nhỏ nhất trong ngôn ngữ. Trong tiếng Việt, một âm tiết bao giờ cũng được phát ra với một thanh điệu, và tách rời với âm tiết khác bằng một khoảng trống (space). Trên chữ viết, mỗi âm tiết tiếng Việt được ghi thành một " chữ " . Vd: "hoa hồng bạch" gồm 3 chữ hoặc 3 âm tiết. .

2. **Con chữ nguyên âm** là những con chữ thường dùng để viết các nguyên âm:

a, ă, â, e, ê, i, o, ô, ơ, u, ư, y.

Chú ý: Trong tiếng Việt, các con chữ O và U được dùng không chỉ để viết các nguyên âm o và u, mà còn để viết một bán nguyên âm (hay bán phụ âm) w (đọc là uờ). Khi dùng O và U để viết w đóng vai trò là âm đệm trong các trường hợp như *toán, toàn, xoan, tuần, tuấn, quần...* thì gọi là *bán nguyên âm*. Khi dùng O và U để viết w đóng vai trò là âm cuối trong các trường hợp như *đào hào, báo cáo, táo, đau, rau câu...* thì gọi là *bán phụ âm*. Với con chữ I, cũng tương tự, nó vừa dùng để viết nguyên âm i (im, in, ít...), vừa dùng để viết bán phụ âm i đóng vai trò là âm cuối trong các trường hợp như: *tai tái, cày cấy, tấy máy...* (xem thêm *Bảng nguyên âm*). Chúng tôi gộp chung tất cả vào nhóm *con chữ nguyên âm*.

3. **Con chữ phụ âm** là những con chữ chỉ dùng để viết các phụ âm:

b, c, d, đ, (f), g, h, (j), k, l, m, n, p, q, r, s, t, v, (w), x, (z).

4. **Tổ hợp con chữ phụ âm** là những tổ hợp hai hoặc ba con chữ phụ âm, dùng để viết một phụ âm:

ch, gh, kh, ng, ngh, nh, ph, th, tr, gi, qu.

Chú ý: Trường hợp **gi** và **qu** thì u và i không dùng để viết một nguyên âm nào cả, cũng gọi là những *tổ hợp con chữ phụ âm*.

#### QUY TẮC

**Quy tắc 1.** Với những âm tiết chỉ có một con chữ nguyên âm, thì dấu thanh được đặt vào con chữ nguyên âm đó. Ví dụ:

á à, ì ách, ọ ẹ, ử rữ, ọp ẹp, ục ịch, hà, lán, giá, giục, quả, quỹ, quýt...  
(trường hợp gi và qu xem **định nghĩa 4**)

**Quy tắc 2.** Với những âm tiết, mà trong âm tiết đó chỉ cần có **một con chữ nguyên âm mang dấu phụ** (Ă, Â, Ê, Ô, Ơ, U) và không kể kết thúc bằng con chữ gì, thì dấu thanh bao giờ cũng đặt ở con chữ đó (riêng ƯƠ, dấu đặt ở Ơ):

ế ẩ, ố ẻ, ở ẻ, ứ ừ, chiến chiến, cuội, cừu, duệ, duệnh, giới, giường, ngoẵng, quyết, ruỗng, rựu, siết, suyển, tuần tiết, tiến triển...

**Quy tắc 3.** Với những âm tiết có **hai con chữ nguyên âm** và kết thúc bằng một con chữ phụ âm hoặc

NHẬN DẠNG THỰC THỂ  
ĐỊNH DANH TRONG VĂN  
BẢN TIẾNG VIỆT

XÂY DỰNG MÔ HÌNH NGŨ  
NGHĨA ÁP DỤNG CHO VIỆC  
BIÊN SOẠN TỪ ĐIỂN GIẢI  
THÍCH

Hướng tới việc xây dựng  
MẠNG TỪ tiếng Việt

Xây dựng kho ngữ liệu áp  
dụng cho phân tích, xử lý  
ngôn ngữ và biên soạn từ  
điển

Ứng dụng phương pháp  
Pointwise vào bài toán tách  
từ cho tiếng Việt

Xác định dãy từ đồng nghĩa  
khi xây dựng Wordnet tiếng  
Việt

Xác định ĐƠN VỊ TỪ VỰNG  
MỚI xuất hiện trong văn bản  
tiếng Việt

Ngôn ngữ học máy tính và  
việc xây dựng từ điển

tổ hợp con chữ phụ âm, thì dấu thanh được đặt vào con chữ nguyên âm chót. Ví dụ:

*choàng, hoạch, loét, quệt, suýt, thoát, xoè/xoét...*

**Quy tắc 4.** Với những âm tiết kết thúc bằng **oa, oe, uy**, dấu thanh được đặt vào con chữ nguyên âm chót. Ví dụ:

*hoạ, hoè, huỷ, loà/xoà, loé, suý, thuỷ...*

**Quy tắc 5.** Với những âm tiết kết thúc bằng **hai** hay **ba con chữ nguyên âm** khác với **oa, oe, uy**, thì dấu thanh được đặt vào con chữ nguyên âm áp chót. Ví dụ:

*bài, bẫy, chĩa, chịu, cửa, đào hào, giúi, hoại, mía, ngoáy, ngoáo, quau, queo, ngoẻo, chịu, chĩa...*

### Chú ý

Có một vấn đề được nhiều người quan tâm trong thời gian gần đây, đó là vị trí dấu thanh trong các văn bản, sách báo tiếng Việt thường không xử lý thống nhất với nhau, gây không ít khó khăn cho việc trao đổi cũng như học tập tiếng Việt. Trước nay người ta chỉ quan tâm đến việc đặt dấu thanh vào đâu trong âm tiết cốt để tạo nên "một cái nhìn" cân đối cho chữ viết. Nhưng ngày nay, với sự phát triển mạnh mẽ của tin học, máy tính ngày càng thâm nhập rộng rãi trong mọi lĩnh vực của đời sống xã hội thì yêu cầu thống nhất trong lưu trữ, xử lý cũng như trao đổi ngữ liệu là quan trọng hơn nhiều so với việc xem chữ viết bỏ dấu có cân đối hay không. Cần khẳng định lại mấy điểm sau đây:

**1.** Chuẩn về vị trí dấu thanh trong âm tiết tiếng Việt đã được xác định từ lâu trong phần lớn trường hợp, về cơ bản tuân theo 5 quy tắc nêu trên. Cho đến nay, chỉ tồn tại sự không nhất trí với các tổ hợp *oa, oe, ua, ue, uy*. Chẳng hạn viết HỌA, HỒE, HỦY, QUA, QUÊ, QUÝ, hay HOẠ, HOÈ, HUỖ, QUẢ, QUỀ, QUỖ?

**2.** Trong tiếng Việt, các con chữ O và U được dùng không phải chỉ để viết các nguyên âm o và u, mà còn để viết một bán nguyên âm (còn gọi là bán phụ âm) w trong một số trường hợp nhất định như: *oa* (wa), *oe* (we), *uy* (wi), *qua* (kwa), *que* (kwe), *quy* (kwi) (c, q, k phiên âm quốc tế đều là /k/), v.v... Có nghĩa là âm w trong các tổ hợp nói trên thật ra không thuộc thành phần nguyên âm của âm tiết, nguyên âm ở đây là *a, e, i* (viết bằng Y).

**3.** Theo quy tắc 1, dấu thanh đánh trên con chữ viết nguyên âm; vậy viết HOẠ, HOÈ, HUỖ, QUẢ, QUỀ, QUỖ là hợp quy tắc (giống như viết HẠ, HÈ, KỤ, CẢ, KỂ), nên coi đó là chuẩn, thống nhất với các trường hợp tương tự khác như NGOAN NGOÃN, KHOẢNG ĐẠT, NGÚNG NGUÃY... (không ai viết NGOAN NGỎAN, KHỎANG ĐẠT, NGÚNG NGỦÃY...). Việc bỏ dấu thanh trên O và U trong những tổ hợp này là kết quả của một sự nhầm lẫn, cho rằng O và U ở đây viết các nguyên âm *o* và *u*.

**4.** Các trường hợp trên đây khác các trường hợp *ia, ua, ua*, là những nguyên âm đôi. Đối với nguyên âm đôi, thì dấu thanh đánh trên con chữ đầu của tổ hợp, tức là đánh trên I (IA), U (UA), U' (UA). So sánh CỬA và QUẢ, THỦA và THỎA. Trong *gia* (già/lúa), nguyên âm là a, nên đánh dấu nặng dưới A: GI + A; khác với trong *già* (giắt già), có nguyên âm đôi *ia* (I ở đây vừa tổ hợp với G viết phụ âm *gi*, vừa tổ hợp với A viết nguyên âm đôi *ia*), nên đánh dấu nặng dưới I: GI + IA (đúng ra là có hai chữ I, nhưng đã lược bỏ bớt một thành GIA).

Cũng nên lưu ý là tương ứng với các nguyên âm đôi *ia, ua, ua*, có các nguyên âm đôi *iê-, uô-, ươ-* (luôn luôn có phụ âm hoặc bán nguyên âm cuối), dấu thanh lại đánh trên con chữ thứ hai của tổ hợp IÊ, UÔ, ƯƠ (quy tắc 2). Cho nên những âm tiết như HUỖ, THUỖ, từ lâu viết nhất trí với dấu thanh trên Ê, Ô, không ai đánh dấu thanh trên U và viết HUỖ, THUỖ.

**5.** Đồng thời với việc chuẩn hoá dấu thanh, cũng cần chuẩn hoá và thống nhất chính tả trong các trường hợp sau đây:

a) Nhất loạt viết khuôn vần /-i/ bằng I (trừ tên riêng) trong các âm tiết H-, K-, L-, M-, T- (nhất loạt viết HI, KI, LI, MI, TI, giống như viết BI, CHI, DI, v.v.; không nên viết HY, KY, LY, MY, TY, cũng như không ai viết BI thành BY, CHI thành CHY, v.v.);

b) Nhất loạt viết khuôn vần /-ui/ (u ngắn) bằng UY (nhất loạt viết QUY, giống như viết HUY, NGUY, TUY, v.v.; không nên viết QUI, cũng như không ai viết NGUY thành NGUI, HUY thành HUI, v.v.).

c) Khi "I" đứng một mình làm thành một từ (hoặc một âm tiết), thì: nếu là từ Hán-Việt, nên viết "Y", chẳng hạn viết Y KHOA, Y THỂ, Y KIẾN, ..., không viết I KHOA... Í KIẾN...; nếu là từ thuần Việt, nên viết "I", chẳng hạn viết Í EO, Í ỐI..., không viết Ý EO, Ý ỐI....

**6.** Nếu căn cứ vào lí do thẩm mỹ để cho rằng bỏ dấu theo kiểu HÒA, QUA... là đúng vì nó cân đối, thì trong các trường hợp sau tại sao không bỏ dấu lại cho nó cân đối. Ví dụ:

THUỖ	(thuở ấu thơ)	→	THỬ	(thử ấu thơ)
NGOÃN	(ngoan ngoan)	→	NGỎAN	(ngoan ngoan)
NGOẠN	(ngọan mục)	→	NGỌAN	(ngọan mục)
HUYỄN	(huyền diệu)	→	HUYỀN	(huyền diệu)

v.v.

### Bài đăng trước:

