

## Chương 3

# CÀI ĐẶT CHƯƠNG TRÌNH

### 3.1. HỆ THỐNG CHƯƠNG TRÌNH BẮT LỖI CHÍNH TẢ:

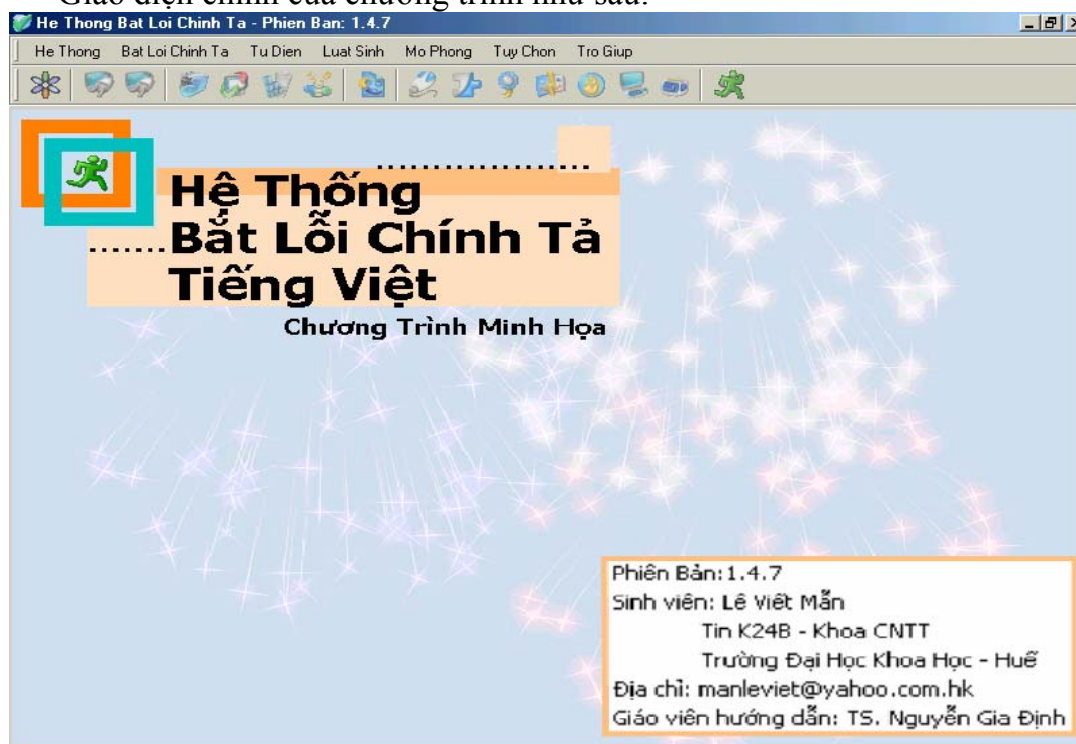
Với tính chất minh họa, khoá luận đã xây dựng chương trình bắt lỗi chính tả và một hệ thống các chương trình hỗ trợ bắt lỗi chính tả. Các chương trình được xây dựng với các ưu điểm sau:

- Vấn đề lưu trữ, hiển thị và xử lý đều được xử lý với mã Unicode.
- Chương trình có giao diện đẹp, dễ sử dụng.
- Hỗ trợ đủ chức năng cần thiết cho người sử dụng.
- Xây dựng theo mô hình hướng đối tượng.

Nhưng chương trình vẫn có một hạn chế lớn: Được cài đặt bằng VB, một ngôn ngữ không hỗ trợ các cấu trúc như cây, bảng băm, map,... và đặc biệt là không hỗ trợ hướng đối tượng hoàn toàn. Do đó, chương trình bị hạn chế về tốc độ.

Do thời gian thực hiện khoá luận là ngắn nên một số vấn đề chưa giải quyết triệt để, như: phân tích văn bản, tối ưu từ điển để có tốc độ tìm kiếm cao nhất (có thể xấp xỉ 58 đơn vị thời gian), cải tiến thuật toán Earley, khả năng học,...

Giao diện chính của chương trình như sau:

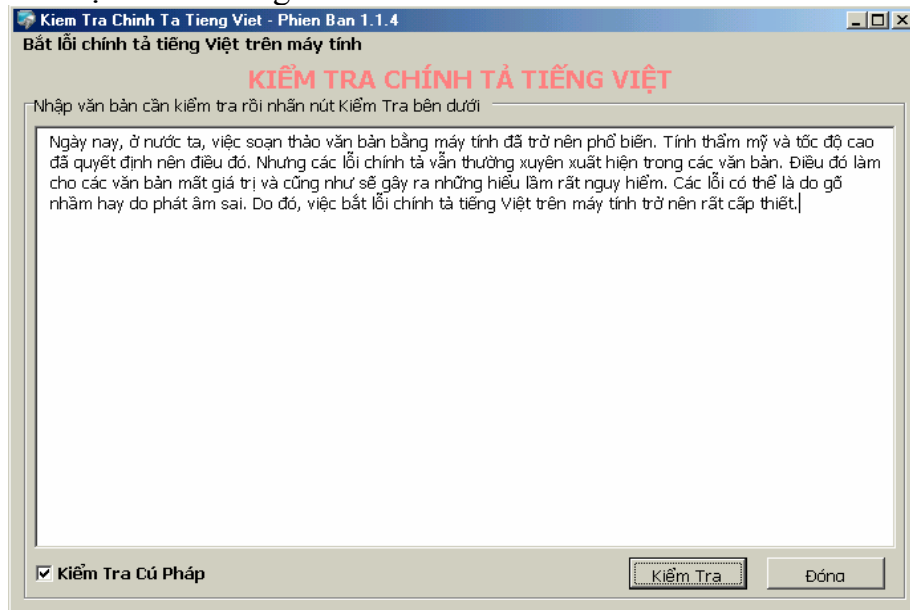


### 3.1.1. Chương trình bắt lỗi chính tả tiếng Việt:

Chương trình bắt lỗi chính tả được xây dựng độc lập (chưa tích hợp vào các chương trình soạn thảo văn bản). Do đó, để có thể bắt lỗi chính tả, người sử dụng phải nhập văn bản vào hoặc sử dụng chức năng Copy\Paste của hệ thống.

Chương trình hỗ trợ cả hai mức bắt lỗi chính tả là mức âm tiết và mức cú pháp. Chương trình mặc định sẽ không kiểm tra mức cú pháp cho văn bản. Nếu muốn kiểm tra mức cú pháp, người sử dụng có thể chọn tùy chọn Kiểm Tra Cú Pháp ở góc dưới cửa sổ chương trình.

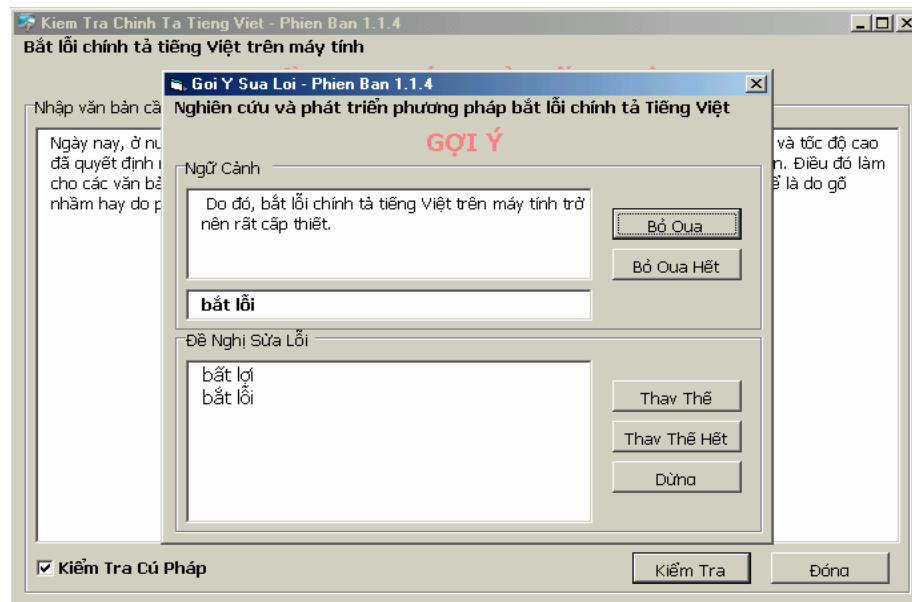
Giao diện của chương trình bắt lỗi chính tả:



Chương trình còn hỗ trợ gợi ý sửa lỗi cho các lỗi mức âm tiết và mức cú pháp. Khi gặp một lỗi, chương trình sẽ thực hiện gợi ý sửa lỗi và hiển thị cửa sổ Gợi Ý. Trên đó, người sử dụng có thể:

- Bỏ qua: Nếu cho rằng từ đó là đúng, kiểm tra từ tiếp theo.
- Bỏ qua hết: Nếu từ đó là hoàn toàn đúng và không muốn chương trình tự động bỏ qua không hỏi tới nếu gặp sau này thì chọn chức năng này.
- Thay thế: Chương trình sẽ thay thế từ sai trên văn bản bằng từ đúng trên cửa sổ Gợi Ý.
- Thay thế hết: Chương trình sẽ thực hiện thay thế tất cả các từ sai giống như vậy bằng từ đúng được chọn.
- Dừng: Chương trình sẽ dừng bắt lỗi.

Giao diện cửa sổ Gợi Ý bên trong Chương trình bắt lỗi chính tả tiếng Việt:



Chương trình sẽ không có chức năng học hay thêm từ mới. Vì đặc điểm của các từ được lưu trong từ điển là bao gồm cả từ loại, việc phân chia từ loại rất phức tạp mà bản thân người sử dụng sẽ gặp nhiều khó khăn, thậm chí là xếp từ loại sai. Do đó, chương trình sẽ không cho người sử dụng thêm từ mới vào. Chức năng này nên được đảm nhiệm bởi đội ngũ phát triển chương trình.

### 3.1.2. Chương trình bắt lỗi chính tả mức âm tiết dựa trên nguyên lý cấu tạo âm tiết:

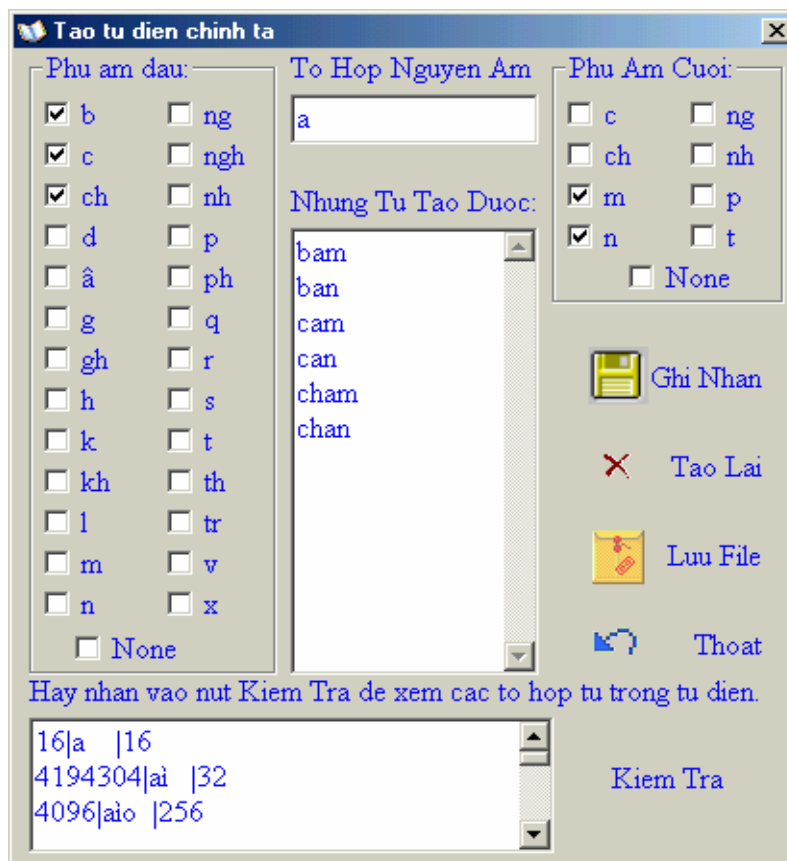
Đây là một chương trình mô phỏng bắt lỗi mức âm tiết dựa trên nguyên lý cấu tạo âm tiết. Do không được sử dụng vào giải thuật bắt lỗi chung nên chương trình được xây dựng riêng và được đặt tên là Chương Trình Kiểm Tra Chính Tả.

Chương trình có ba chức năng:

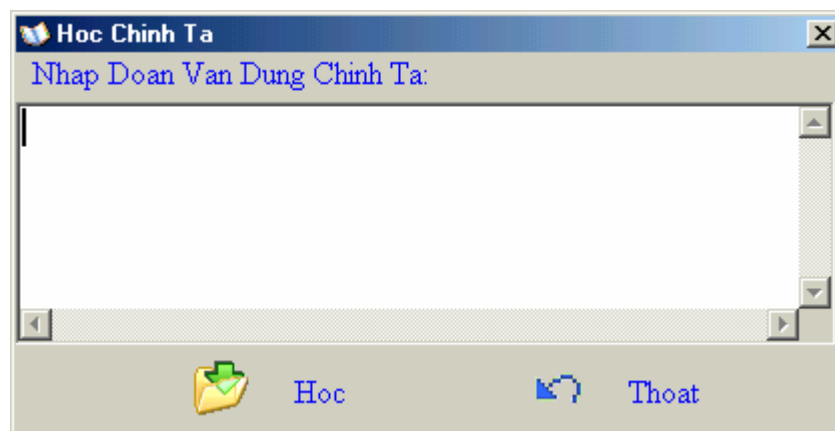
- Tạo từ điển chính tả: Nhập các âm tiết mới vào từ điển.
- Học chính tả: Chương trình tự động học âm tiết.
- Kiểm tra chính tả: Chương trình kiểm tra âm tiết.



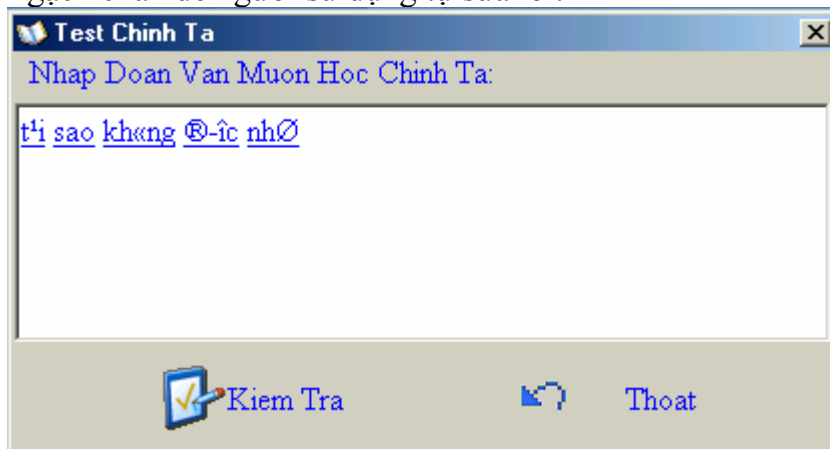
Trong chương trình Tạo từ điển chính tả, sau khi nhập xong tổ hợp các âm tiết cùng nguyên âm rồi thì có thể nhấn nút Hiện Thị để xem các âm tiết có thể tạo ra. Hay nhấn nút Nhập Mới để nhập mới tổ hợp các âm tiết khác. Sau khi xem tất cả các âm tiết có thể tạo ra, người sử dụng có thể nhấn Ghi Nhận để ghi vào mảng các âm tiết hay có thể nhấn Tạo Lại để tạo lại. Nút Lưu File là để lưu từ điển vào file lưu trữ. Nút Thoát là để quay trở lại giao diện chính. Nút Kiểm Tra là để xem tất cả các âm tiết trong từ điển.



Với chương trình Học chính tả, người sử dụng phải nhập một văn bản đúng chính tả rồi nhấn nút Học. Chương trình sẽ tự động phân tích văn bản để học các âm tiết.



Cuối cùng, với chương trình kiểm tra chính tả, người sử dụng cũng phải nhập một văn bản rồi nhấn nút Kiểm Tra. Chương trình sẽ phân tích văn bản ra từng âm tiết một và kiểm tra từng âm tiết xem có đúng không. Các âm tiết sai sẽ được chương trình gạch chân để người sử dụng tự sửa lỗi.

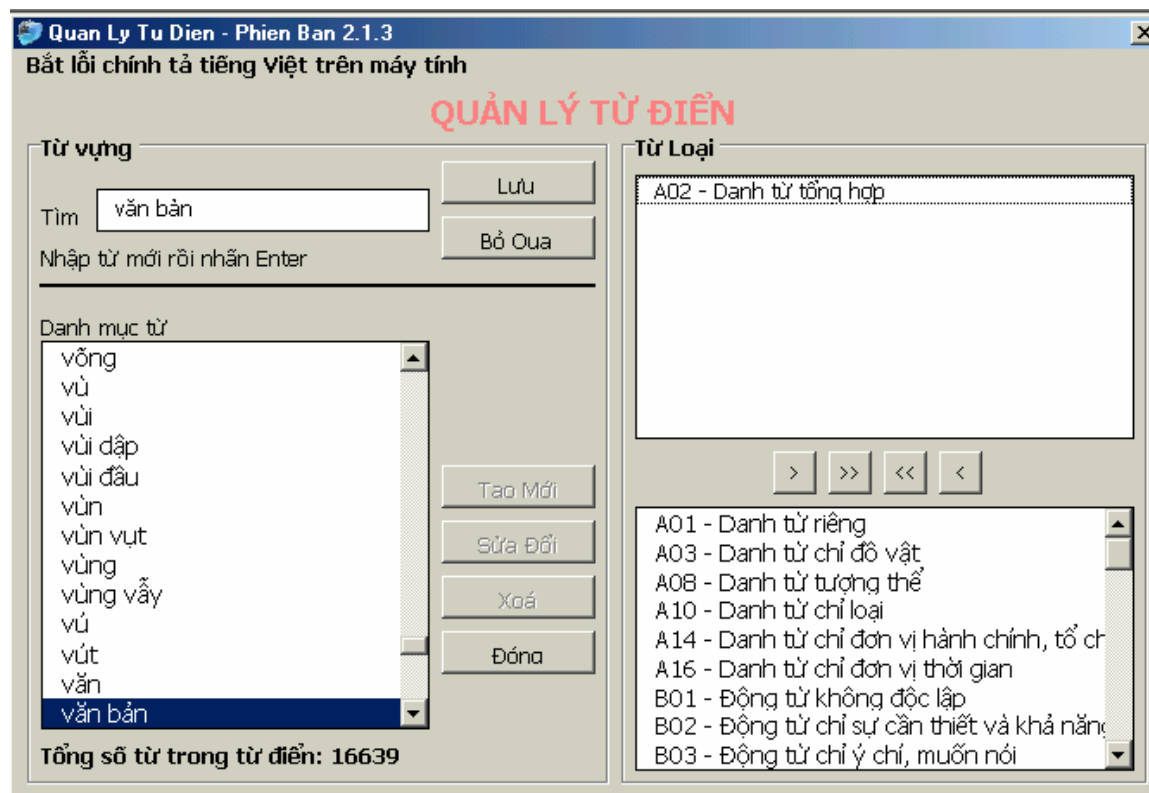


### 3.1.3. Các chương trình hỗ trợ và minh họa:

Sau đây sẽ là một số chương trình hỗ trợ cho chương trình bắt lỗi và vài chương trình minh họa một số giai đoạn trong quá trình phân tích cú pháp.

#### 3.1.3.1. Quản lý từ điển:

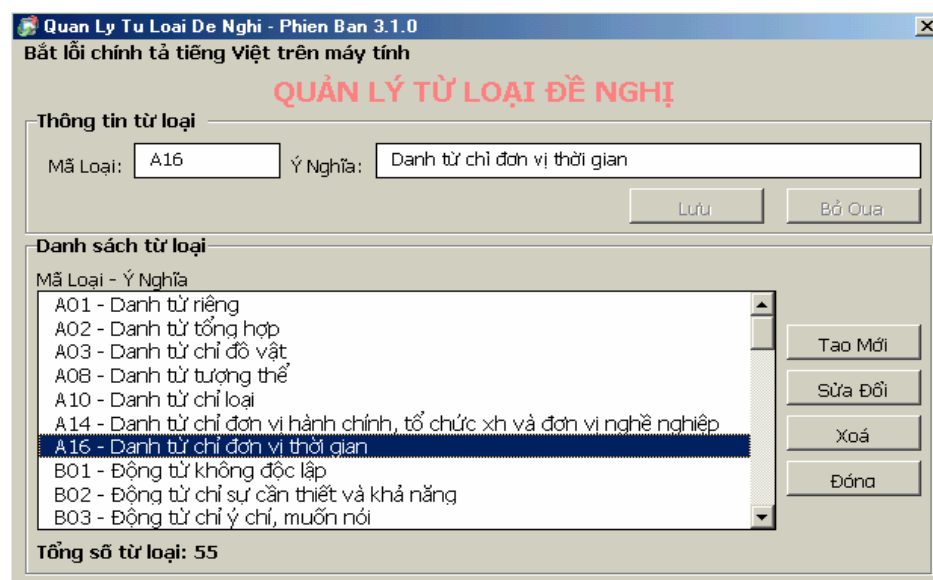
Đây là một trong những chương trình quan trọng hỗ trợ cho chương trình bắt lỗi. Chương trình cho phép thêm, bớt, sửa đổi và xoá các từ khỏi từ điển. Người sử dụng đồng thời có thể tìm kiếm các từ, xem từ loại và sửa đổi từ loại của các từ.



### 3.1.3.2. Quản lý từ loại:

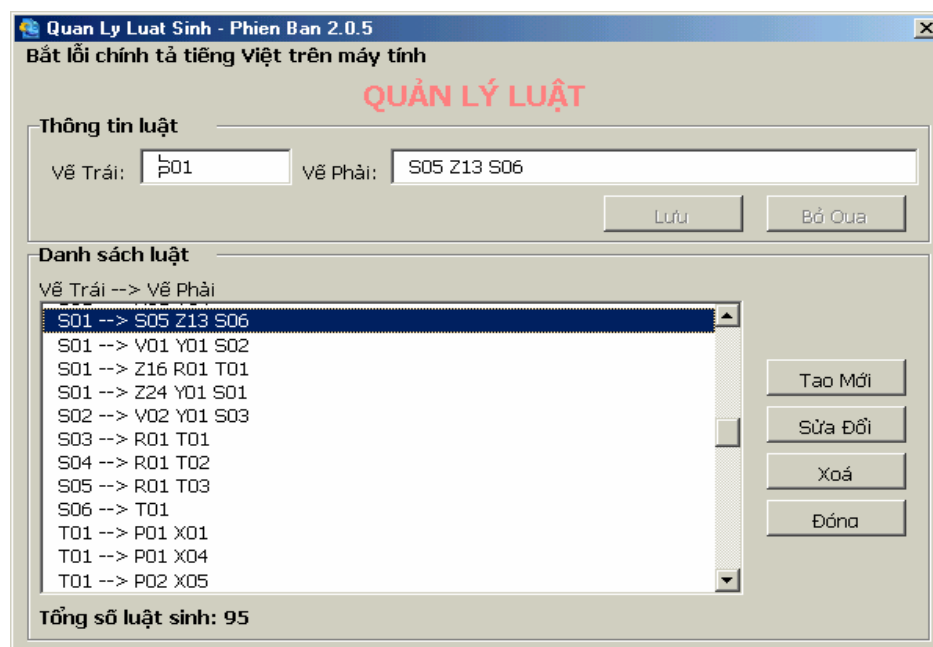
Đây là một chương trình nhỏ, hỗ trợ quản lý các từ loại được lập cho chương trình bắt lỗi. Chương trình có các chức năng sau:

- Tạo Mới: Cho phép nhập mã từ loại mới và ý nghĩa của từ loại đó.
- Sửa đổi: Cho phép sửa đổi mã và ý nghĩa của các từ loại đã có.
- Xoá: Cho phép xoá bỏ các từ loại đã có.



### 3.1.3.3. Quản lý luật sinh:

Để hỗ trợ cho phân tích cú pháp và quản lý bộ luật sinh, khoá luận đã xây dựng chương trình quản lý luật sinh. Chương trình hỗ trợ nhập luật mới, xem các luật đã có, sửa đổi các luật, xoá bỏ luật sai.

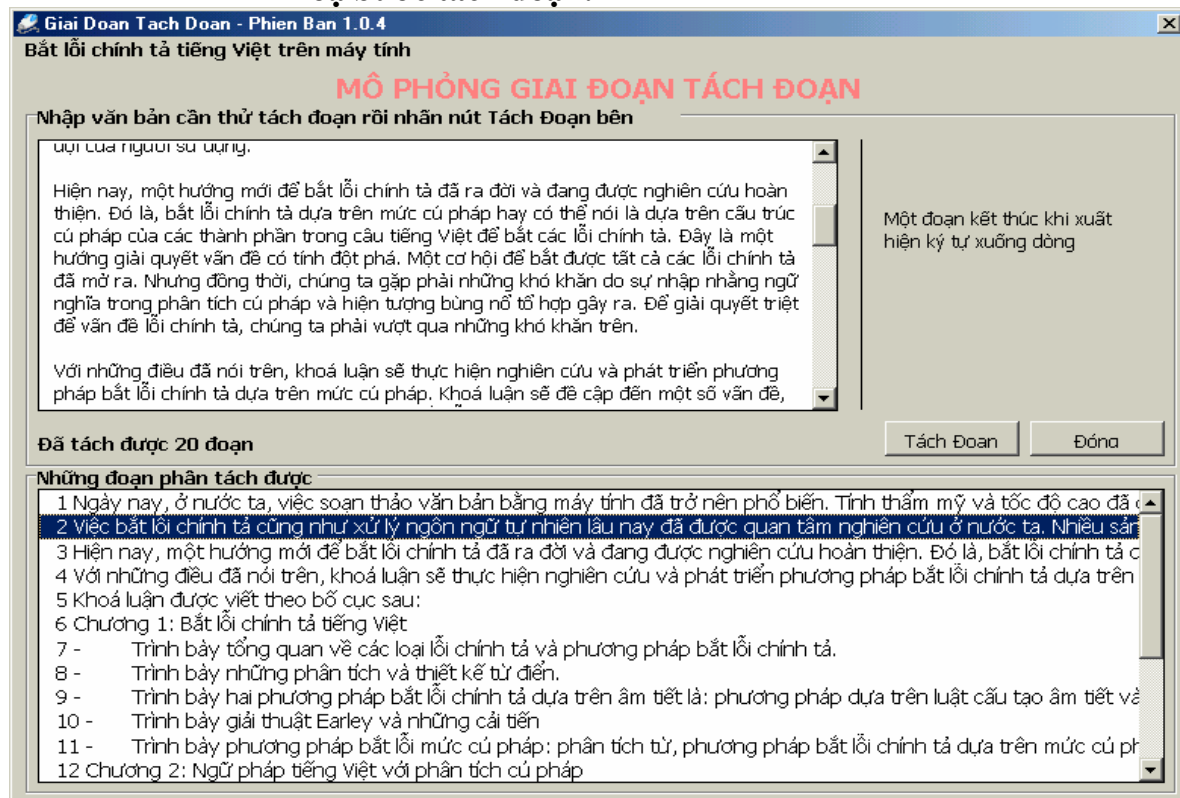




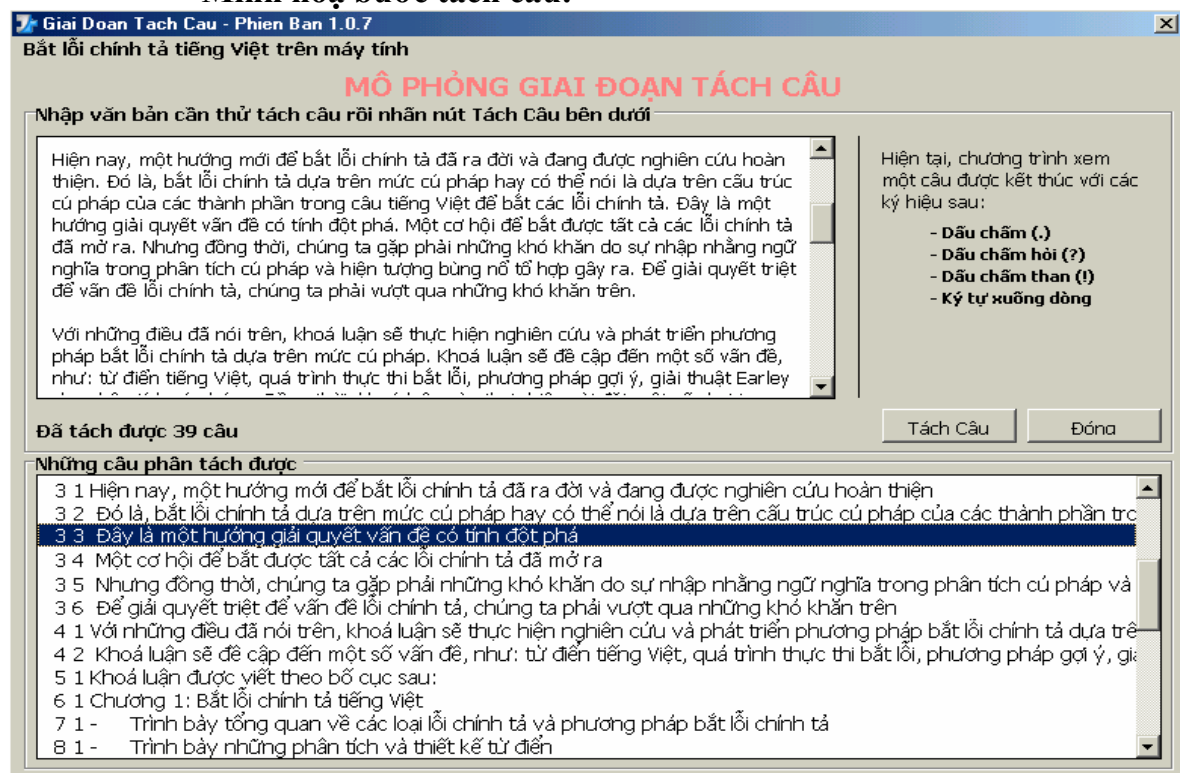
### 3.1.3.4. Các chương trình minh hoạ các giai đoạn phân tích cú pháp:

Để minh hoạ từng bước giai đoạn phân tích cú pháp, khoá luận xây dựng một vài chương trình minh hoạ.

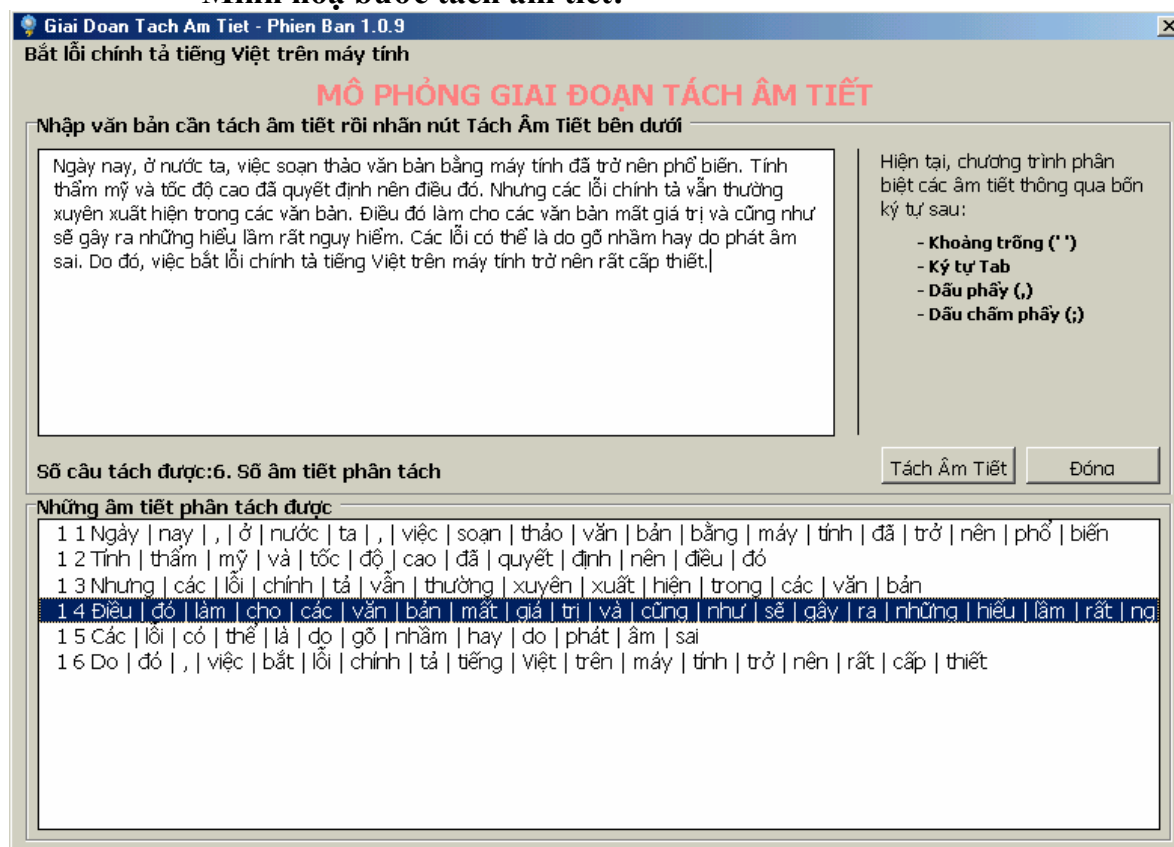
- **Minh hoạ bước tách đoạn:**



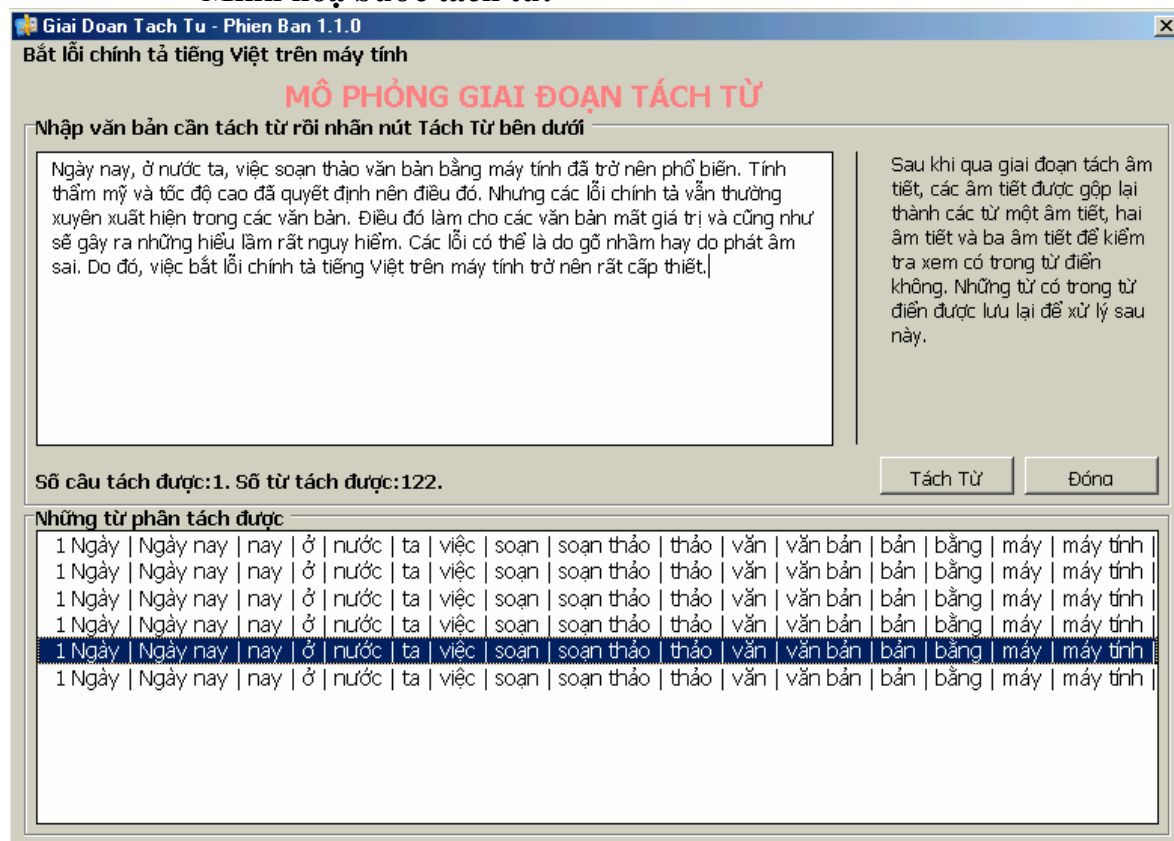
- **Minh hoạ bước tách câu:**



• **Minh hoạ bước tách âm tiết:**

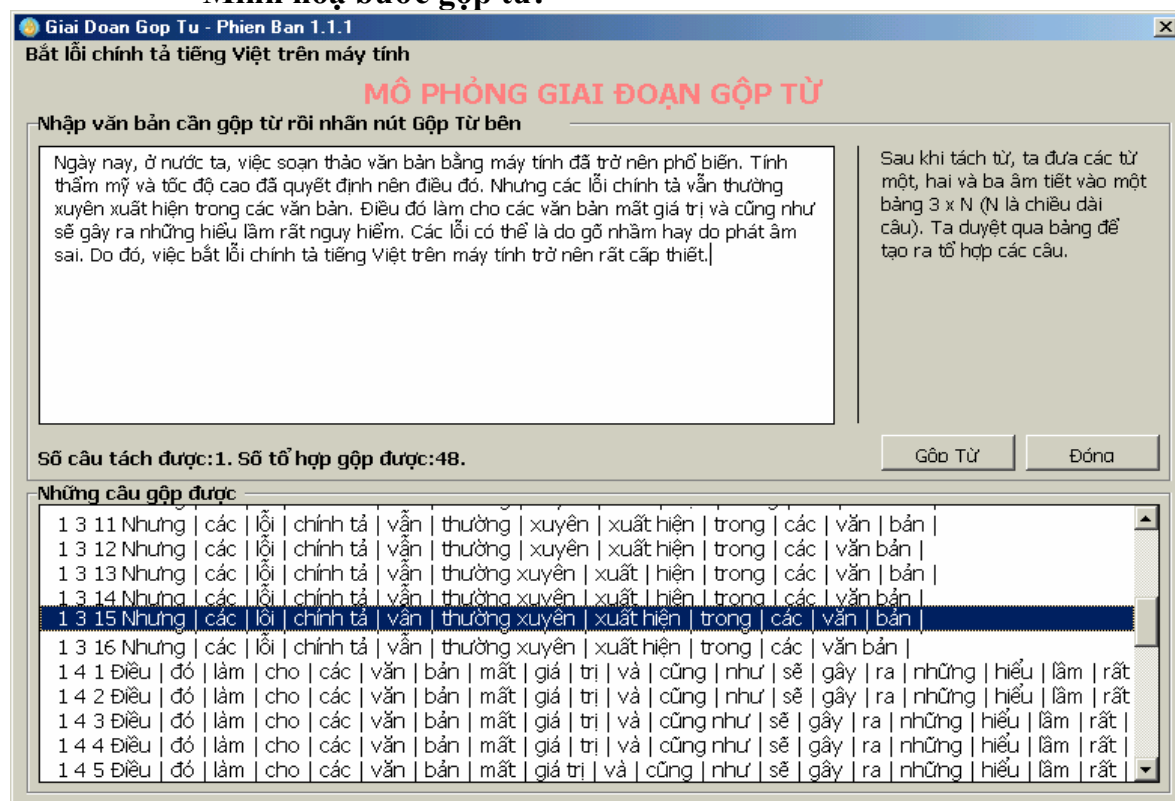


• **Minh hoạ bước tách từ:**

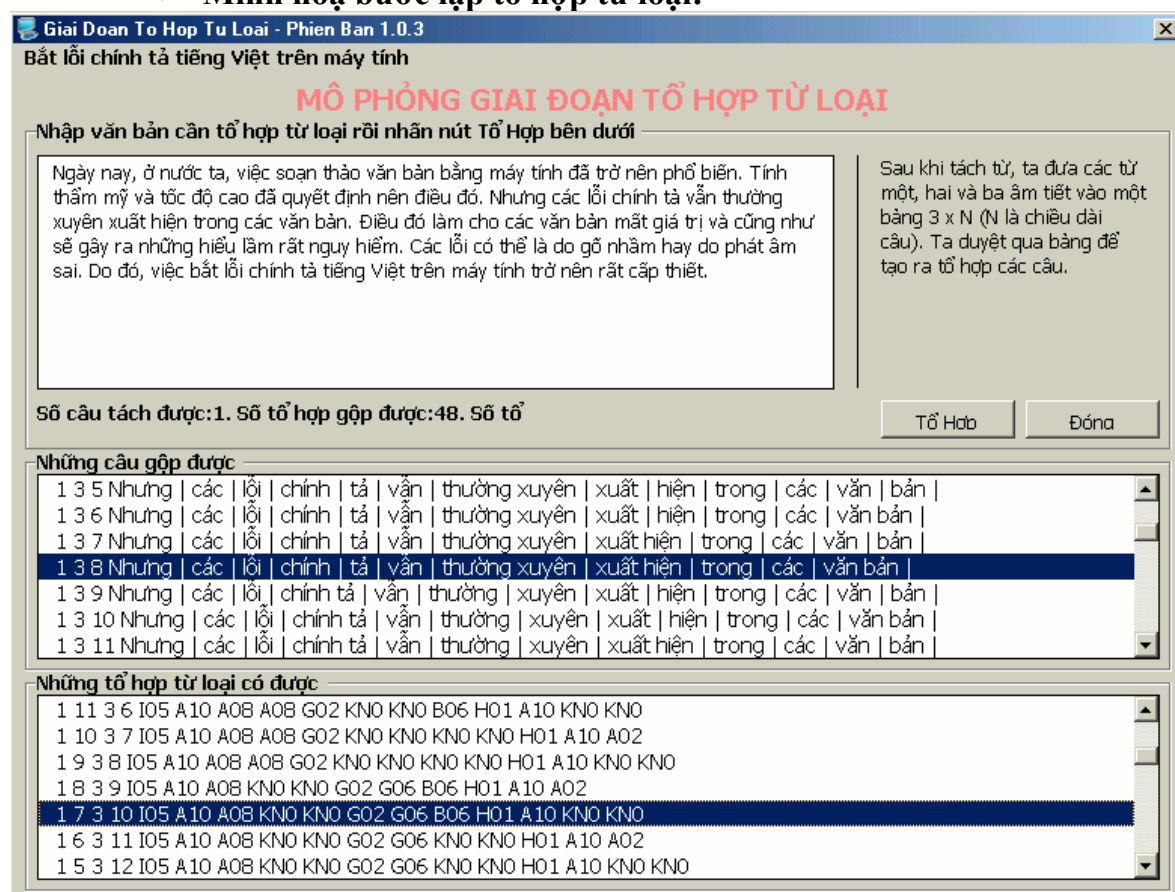




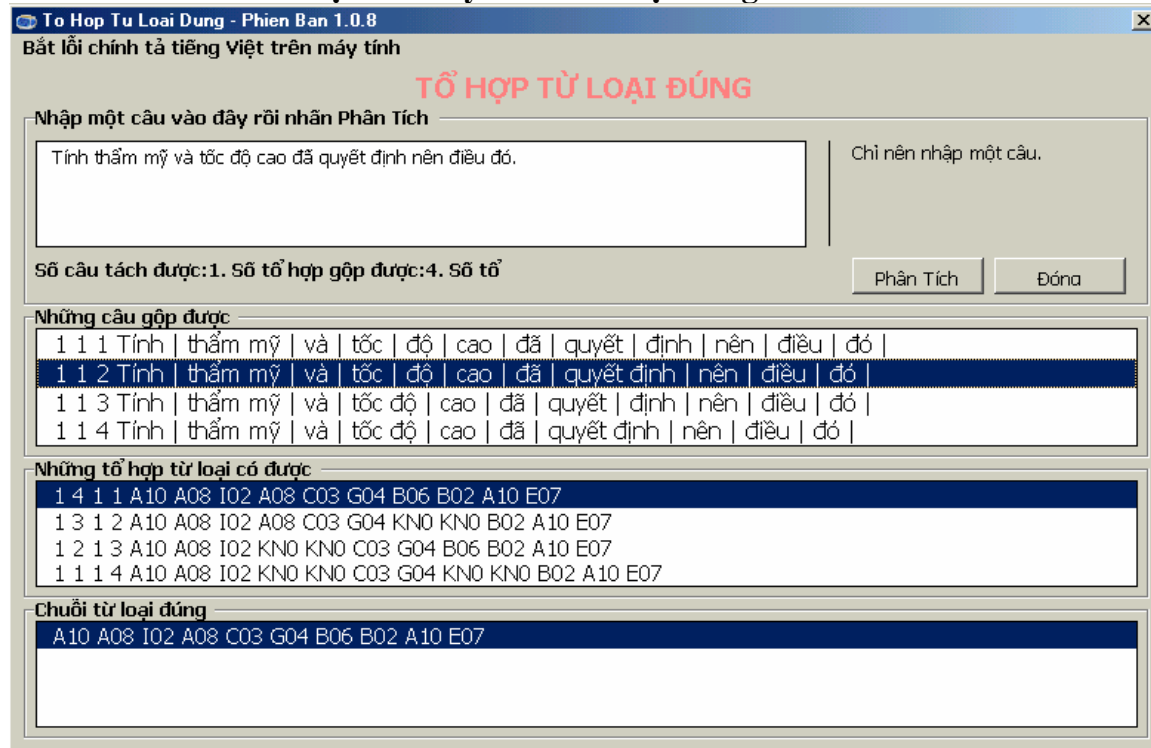
• **Mình hoạ bước gộp từ:**



• **Mình hoạ bước lập tổ hợp từ loại:**



- **Minh hoạ bước lấy chuỗi từ loại đúng:**



### 3.1.3.5. Chương trình học âm tiết và chương trình học từ:

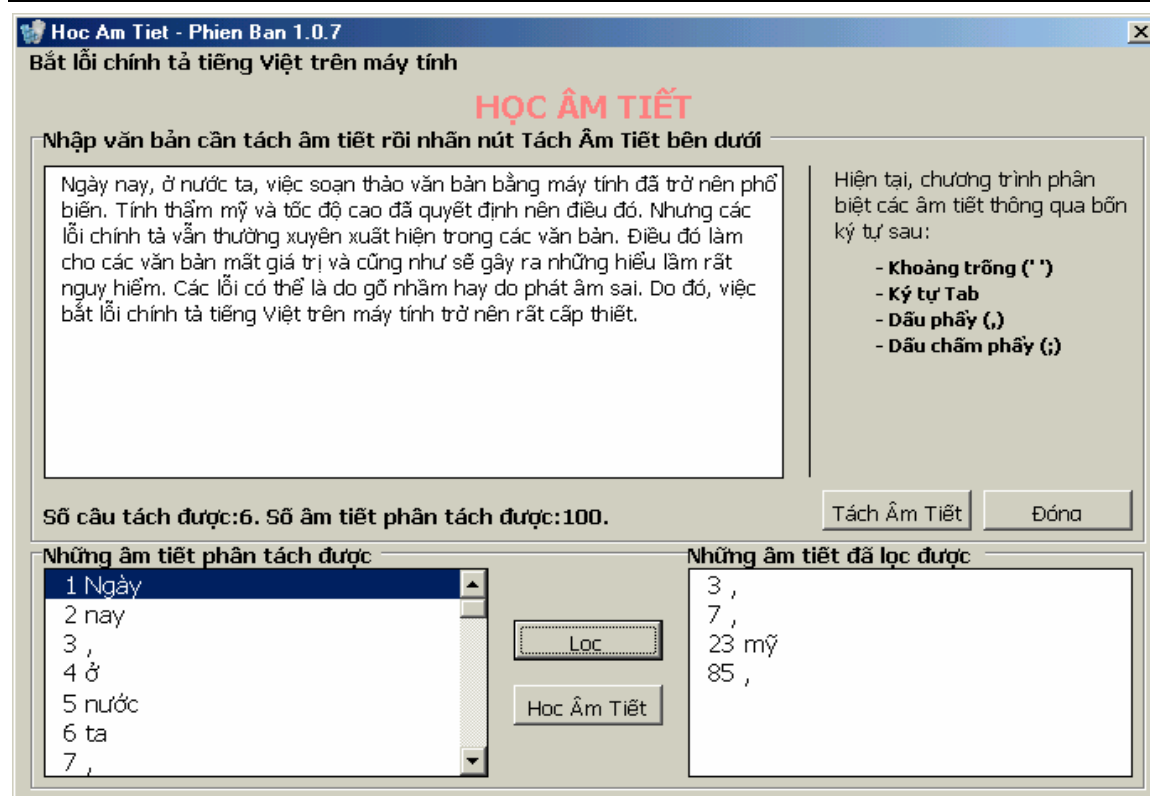
Nhằm tạo thuận tiện trong việc nhập âm tiết và từ mới vào từ điển, khoá luận cũng đã xây dựng hai chương trình tự động phân tích âm tiết và từ để người sử dụng có thể chọn các âm tiết và từ cần nhập vào từ điển.

- **Chương trình học âm tiết:**

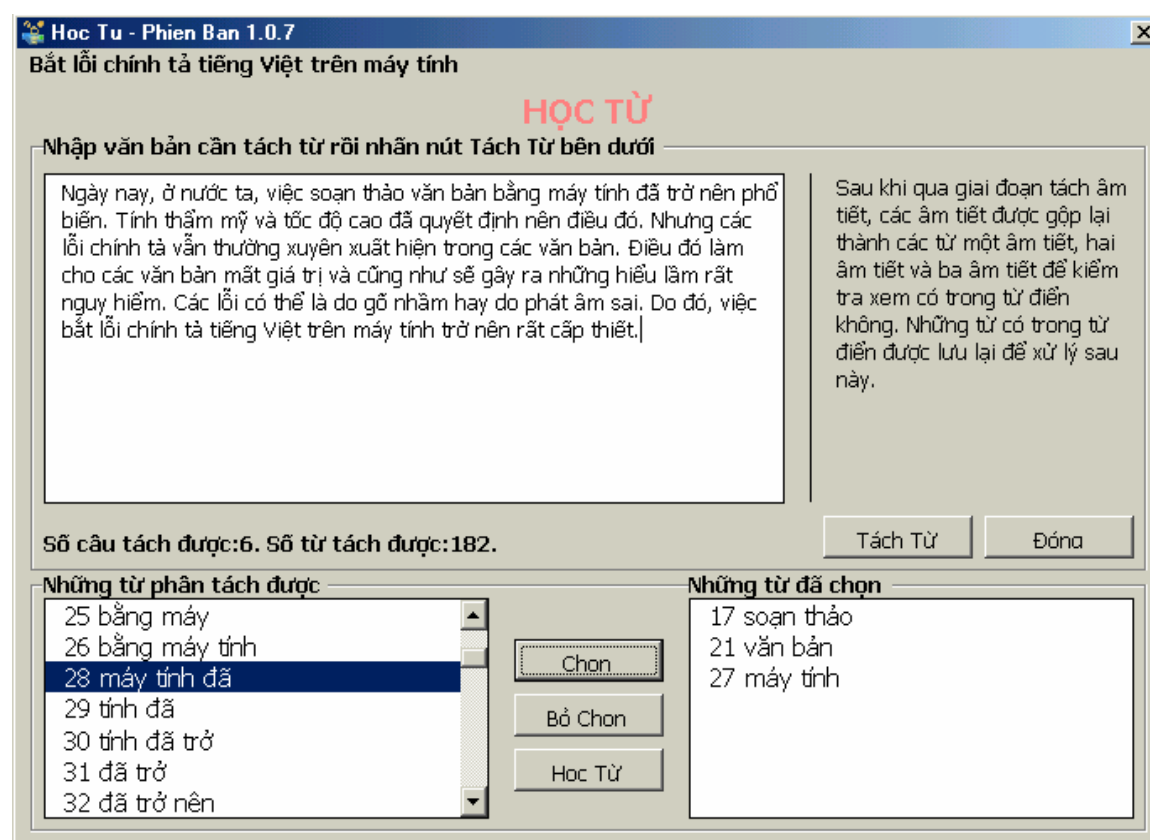
Người sử dụng nhập các âm tiết cách nhau một khoảng trống rồi nhấp nút Phân tích để chương trình tự động tách các âm tiết và liệt kê ở danh sách những âm tiết phân tách được. Đến đây, người sử dụng nhấp nút Lọc để lọc các âm tiết mới. Chương trình tự động lấy lần lượt các âm tiết đã tách được ra kiểm tra xem có trong từ điển hay chưa. Nếu âm tiết chưa có trong từ điển thì sẽ được liệt kê trong danh sách những âm tiết đã lọc được. Người sử dụng có thể nhấp nút Học Âm Tiết để nhập các âm tiết mới vào từ điển.

- **Chương trình học từ:**

Người sử dụng nhập các từ cách nhau một khoảng trống rồi nhấp nút Tách Từ. Chương trình sẽ tách các từ hai âm tiết và ba âm tiết, được liệt kê trong danh sách Những từ phân tách được. Các từ này có thể là đúng cũng có thể sai nên người sử dụng phải chọn từng từ đúng một sang danh sách Những từ đã chọn rồi mới nhấp nút Học Từ để nhập các từ được chọn vào từ điển.



Chương trình học âm tiết



Chương trình học từ

### 3.2. THỬ NGHIỆM:

Tiến hành phân tích đoạn đầu trong Lời nói đầu của khoá luận:

“Ngày nay, ở nước ta, việc soạn thảo văn bản bằng máy tính đã trở nên phổ biến. Tính thẩm mỹ và tốc độ cao đã quyết định nên điều đó. Nhưng các lỗi chính tả vẫn thường xuyên xuất hiện trong các văn bản. Điều đó làm cho các văn bản mất giá trị và cũng như sẽ gây ra những hiểu lầm rất nguy hiểm. Các lỗi có thể là do gõ nhầm hay do phát âm sai. Do đó, việc bắt lỗi chính tả tiếng Việt trên máy tính trở nên rất cấp thiết.”

Đoạn văn trên có tất cả 75 từ với 29 từ loại. Cụ thể như sau:

1. Danh từ chỉ đơn vị thời gian: ngày
2. Danh từ chỉ đơn vị hành chính, tổ chức xã hội và đơn vị nghề nghiệp: nước
3. Đại từ chỉ thời gian: nay
4. Đại từ nhân xưng: ta
5. Giới từ: ở, bằng, trên, trong, như
6. Danh từ chỉ loại: việc, cái, tính, điều, tiếng
7. Danh từ tổng hợp: văn bản
8. Danh từ chỉ đồ vật: máy tính
9. Danh từ trừu tượng: thẩm mỹ, tốc độ, lỗi, giá trị, chính tả
10. Tính từ chỉ lượng: cao
11. Phó từ chỉ quan hệ thời gian đứng trước: đã, sẽ
12. Động từ chỉ sự bắt đầu, tiếp diễn, sự chấm dứt: trở nên
13. Tính từ chỉ quan hệ: nguy hiểm
14. Liên từ liên hợp: và
15. Động từ chỉ sự cần thiết và khả năng: nên, có thể
16. Đại từ chỉ định: đó
17. Động từ chỉ hoạt động: quyết định, gõ, phát âm, bắt, làm, soạn thảo, xuất hiện, mất, gây, hiểu
18. Tính từ chỉ phẩm chất: cấp thiết, sai, nhầm, lầm, phổ biến
19. Phó từ chỉ mức độ đứng trước: rất
20. Quán từ: những, các
21. Trợ động từ: là
22. Liên từ lựa chọn: hay
23. Liên từ kéo theo: do đó, do
24. Danh từ riêng: Việt
25. Phó từ chỉ hướng: ra
26. Phó từ chỉ sự tiếp diễn, sự tương tự: cũng, vẫn
27. Phó từ chỉ tần số: thường xuyên
28. Phó từ cho: cho
29. Liên từ tương phản: nhưng

Các luật sinh được sử dụng cho đoạn văn trên là:

- <Câu> → <Liên từ> <Chủ ngữ> <Vị ngữ>  
 <Câu> → <Trạng ngữ>, <Chủ ngữ> <Vị ngữ>  
 <Câu> → <Chủ ngữ> <Vị ngữ>

nhưng..., do đó...

<Câu> → <Câu> <Liên từ> <Câu>	... và ...
<Câu> → <Vị ngữ>	
<Trạng ngữ> → <Trạng ngữ>, <Trạng ngữ>	
<Trạng ngữ> → <Ngữ danh từ>	ngày nay
<Trạng ngữ> → <Giới từ> <Ngữ danh từ>	ở nước ta
<Chủ ngữ> → <Danh từ chỉ loại> <Ngữ danh từ>	các lỗi chính tả
<Chủ ngữ> → <Ngữ danh từ> <Liên từ> <Ngữ danh từ>	... và ...
<Chủ ngữ> → <Ngữ danh từ>	điều đó
<Ngữ danh từ> → <Danh từ> <Ngữ danh từ>	lỗi chính tả
<Ngữ danh từ> → <Danh từ chỉ loại> <Danh từ>	các văn bản
<Ngữ danh từ> → <Danh từ>	
<Ngữ danh từ> → <Động từ> <Ngữ danh từ>	soạn thảo...
<Ngữ danh từ> → <Ngữ danh từ> <Định từ>	
<Ngữ danh từ> → <Danh từ> <Tính từ>	tốc độ cao
<Ngữ danh từ> → <Quán từ> <Ngữ danh từ>	những hiểu lầm
<Ngữ danh từ> → <Danh từ> <Đại từ>	
<Định từ> → <Giới từ> <Danh từ>	bằng máy tính
<Vị ngữ> → <Ngữ động từ> <Bổ ngữ>	
<Ngữ động từ> → <Phó từ> <Ngữ động từ>	vẫn thường xuyên xuất hiện
<Ngữ động từ> → <Động từ> <Động từ>	quyết định nên, làm cho
<Ngữ động từ> → <Động từ> <Tính từ>	gõ nhầm, phát âm sai
<Ngữ động từ> → <Động từ>	trở nên
<Bổ ngữ> → <Giới từ> <Ngữ danh từ>	trong các văn bản
<Bổ ngữ> → <Tính từ>	
<Bổ ngữ> → <Ngữ danh từ>	điều đó
<Bổ ngữ> → <Ngữ danh từ> <Bổ ngữ>	
<Bổ ngữ> → <Động từ> <Ngữ danh từ>	mất giá trị
<Bổ ngữ> → <Phó từ> <Bổ ngữ>	rất nguy hiểm

Với phân tích về từ loại và luật sinh như trên, chương trình bắt lỗi sẽ kiểm tra và báo đúng cho đoạn văn trên. Những câu không phù hợp về cú pháp sẽ bị chương trình bắt lỗi và sẽ đưa ra gợi ý sửa lỗi.