

Testmethoden für die sichere, sinnentsprechende Silbentrennung und andere Anwendungen einer Wortanalyse

Martin Schönhacker*, Gabriele Kodydek,
Institut für Computergraphik, Abteilung für Algorithmen und Datenstrukturen,
Technische Universität Wien, Favoritenstraße 9-11/186, A-1040 Wien, Österreich
{schoenhacker,kodydek}@apm.tuwien.ac.at, <http://www.apm.tuwien.ac.at/>

Kurzfassung

Es wird eine Testumgebung für die sichere, sinnentsprechende Silbentrennung (*SiSiSi*) deutscher Wörter vorgestellt, die auf einem Wortanalyse-System aufbaut.¹ Eine wesentliche Aufgabe beim Trennen von Wörtern ist die Erkennung und Aufspaltung zusammengesetzter Wörter, die im Deutschen in unbegrenzter Zahl gebildet werden können. Durch diese Wortbildungsart ist es nicht möglich, das System mit Hilfe einer Liste aller existierenden Wörter und ihrer korrekten Zerlegungen zu testen. Daher wurden spezielle Testmethoden für große Textdateien entwickelt, welche die wenigen problematischen Fälle aus der großen Menge analysierter Wörter herausfiltern. Diese Methoden beinhalten die Erkennung unbekannter oder mehrdeutig zerlegbarer Wörter, den Vergleich der Resultate verschiedener Versionen des Wortanalyse- und Silbentrennsystems, und die genauere Inspektion „verdächtiger“ Wörter mit Hilfe von Spezialkriterien. Die Testumgebung ist auch für andere auf Wortanalyse basierende Anwendungen geeignet, wie zum Beispiel für sinnentsprechende Volltextsuche.

1 Einleitung

Im vorliegenden Artikel werden Methoden zum Testen der sicheren und sinnentsprechenden Silbentrennung deutschsprachiger Wörter vorgeschlagen, welche auf ein Wortanalyse-System aufbaut. Ein spezielles Ziel der vorgestellten Testumgebung war es, den Vergleich der Resultate verschiedener Versionen der Wortanalyse zu ermöglichen.

In Abschnitt 2 folgt eine kurze Beschreibung der Silbentrennmethode *SiSiSi* (*Sichere Sinnentsprechende Silbentrennung*). Abschnitt 3 beschreibt die Testumgebung. Die Resultate empirischer Tests folgen in Abschnitt 4. Den Abschluss bildet in Abschnitt 5 ein Ausblick auf die Möglichkeiten und Ideen, andere Anwendungen in das Testsystem aufzunehmen, gefolgt von der Zusammenfassung in Abschnitt 6.

2 Sichere, sinnentsprechende Silbentrennung

Die Methode *SiSiSi* bietet sichere, sinnentsprechende Silbentrennung für deutschsprachige Wörter [1], siehe auch [3]. Weil das komplizierteste Problem in diesem

Zusammenhang die Trennung zusammengesetzter Wörter darstellt, ist die Grundlage des Verfahrens die Wortanalyse: Zusammengesetzte Wörter werden in Einzelwörter unterteilt, welche wiederum nach den Regeln der Wortgrammatik in atomare Bestandteile (*Atome*) zerlegt werden.

Nach ihrer Verwendung in der Wortsynthese können Atome grob als Vorsilben, Stämme und Endungen klassifiziert werden. In der Praxis erweist sich allerdings eine detailliertere Klassifizierung der Atome als vorteilhaft, insbesondere wegen der Möglichkeit, danach die Menge der für bestimmte Wortklassen verwendbaren Endungen einzuschränken [6, 7].

Das Wortanalyse-System besteht aus zwei grundlegenden Bestandteilen: die *Atomtabelle* enthält alle dem System bekannten atomaren Wortbestandteile; der rekursive *Zerlegungsalgorithmus* führt die eigentliche Wortanalyse durch, indem er Wörter nach den Regeln der Wortbildungsgrammatik zerlegt.

Nach der Analyse sind Trennstellen an den Wortfugen zusammengesetzter Wörter sowie zwischen Stämmen und Vorsilben bzw. vor bestimmten Endsilben bereits bekannt. Danach wird ein weiterer Algorithmus eingesetzt, der basierend auf der Abfolge von Konsonanten und Vokalen weitere Trennstellen in den identifizier-

* momentan: Franklin & Marshall College, Lancaster, PA 17604–3003, USA, m_schoenhack@acad.fandm.edu

¹ Dieses Projekt wurde zum Teil durch die freundliche Unterstützung der Hochschuljubiläumsstiftung der Stadt Wien unter Geschäftszahl H–75/99 ermöglicht, für die sich die Autoren herzlich bedanken.

ten Stamm-Endungs-Kombinationen und in mehrsilbigen Vorsilben (z.B. *unter*) findet.

Für jedes Eingabewort ist das Resultat des Trennverfahrens eine Menge (im mathematischen Sinn) von Trennvarianten für dieses Wort. Trennstellen zwischen Einzelwörtern einer Zusammensetzung werden als *Haupttrennstellen* (in der Anzeige „=“) markiert, Trennstellen innerhalb von Einzelwörtern als *Nebentrennstellen* (dargestellt durch „-“). Manche Nebentrennstellen, insbesondere jene knapp nach einer Haupttrennstelle, werden als wenig empfehlenswert markiert (und durch „_“ dargestellt). Das Wort *Wort=zer_le-gungs=ver_fah-ren* ist ein gutes Beispiel für den Einsatz dieser unterschiedlichen Trennstellentypen.

Verschiedene Varianten der Zerlegung eines bestimmten Wortes resultieren oft aus der Tatsache, dass es mehrere grammatikalisch korrekte Zerlegungen gibt, die aber nicht notwendigerweise alle sinnvolle Wörter ergeben müssen. Ein Beispiel ist das Wort *Per-so-nal=man-gel* bzw. *Per-son=alm=an_gel*, wo die zweite Bedeutungs- und damit Trennvariante offensichtlich nicht sehr sinnvoll ist. Nur Trennstellen, die in allen Zerlegungsvarianten vorkommen, werden als sicher betrachtet und in das Endresultat aufgenommen (also im Beispiel: *Per-sonalman_gel*).

3 Das Testsystem

3.1 Grundlegende Funktionsweise

Es wird eine große Zahl von Tests benötigt, um sicherzustellen, dass kein richtig geschriebenes Wort eine falsche Worttrennung erhält, und dass orthographisch und grammatikalisch illegale Wörter als solche erkannt werden. Weil SiSiSi nach allen erlaubten Zerlegungen eines Wortes sucht, gilt folgendes: Wenn die Atomtabelle alle atomaren Bestandteile der deutschen Sprache enthält und die verwendeten Wortbildungsregeln keine korrekte Variante der Wortbildung ausschließen, kann für jedes korrekt geschriebene Wort keine der gefundenen *sicheren Trennstellen* (das sind jene, die in allen Trennvarianten vorkommen) falsch sein. Das ergibt sich unmittelbar aus der Tatsache, dass die Menge gefundener Zerlegungen jedenfalls die korrekte Zerlegung enthält.

Im Idealfall sollte jedes existierende Wort geprüft und kontrolliert werden. Leider ist diese Variante nicht gangbar, weil in der deutschen Sprache eine riesige Anzahl von neuen zusammengesetzten Wörtern gebildet werden kann. Daher werden selektive Tests eingesetzt, welche nur Wörter, die bestimmten Kriterien entsprechen, aus der Gesamtmenge herausfiltern.

Es wurde ein Testsystem implementiert, das Wortanalyse und Worttrennung prüft. Das System enthält drei

verschiedene Versionen von SiSiSi. Zwei dieser Versionen verwenden eine sehr detaillierte Klassifikation von Atomen und eine Grammatik, die sich aus den Regeln zur deutschen Wortbildung ableitet. Die erste (*Resi* genannt) folgt den Regeln der Neuen Rechtschreibung, welche seit 1998 in Kraft sind; die zweite (*HelSi*) folgt den „alten“ Regeln, welche übergangsweise noch bis 2005 gelten. Zu Vergleichszwecken wurde auch eine frühere, relativ primitive Version (*UrSi*) aufgenommen, welche eine extrem vereinfachte Wortgrammatik nach den alten Rechtschreibregeln enthält; diese Version erlaubt allerdings die Erkennung einer hohen Zahl sinnloser Wortzusammensetzungen, zum Beispiel durch wiederholtes Anhängen der gleichen Endung an einen Stamm (was dann etwa *Be-haus-ung-en-en-en...* legal macht).

Das hier vorgestellte Testsystem erlaubt die einfache Überprüfung der auf Wortanalyse basierenden Silbentrennung. Sowohl Einzelwörter als auch ganze Textdateien in verschiedenen Formaten (insbesondere mit verschiedenen Darstellungen der deutschen Umlaute und anderer Sonderzeichen wie ß oder é) können geprüft werden. Die Resultate der Wortzerlegung werden in strukturierter Form am Bildschirm angezeigt und können auch in einer Datei mitprotokolliert werden. Es stehen verschiedene Ansichten zur Verfügung, um tieferen Einblick in die Funktionsweise des Algorithmus zu erhalten.

Als Beispiel wollen wir das Wort *Wachstuben* verwenden. Es kann auf zwei Arten gebildet werden, die beide erlaubt und sinnvoll sind, aber in verschiedenen Trennungsvarianten resultieren: entweder aus den Atomen *wach* (der Stamm des Verbs *wachen*) und *stuben* (Räume), oder aus *wachs* (ein Bienenprodukt) und *tuben* (wie bei Zahnpasta). Hierbei werden die Wörter *stuben* und *tuben* automatisch in ihren Stamm (*stub* bzw. *tub*), das erweiternde *e* und die Pluralendung *n* zerlegt. **Bild 3.1** zeigt einen Auszug aus den Resultaten für dieses Beispiel.

Trennvektoren:	
+ wach=stu-ben	
+ wachs=tu-ben	
= wachstu-ben	<i>kombinierter Trennvektor</i>
Atomfolgen:	
+ wach·stub·e·n	
+ wachs·tub·e·n	

Bild 3.1 Trennvektoren und Atomfolgen für das zusammengesetzte Wort *Wachstuben*

Die *Trennvektoren* genannte Ansicht zeigt alle möglichen Arten, das Wort zu trennen, und darunter den sogenannten kombinierten Trennvektor, welcher durch die Kombination aller Trennvarianten entsteht. Dieser enthält nur die sicheren Trennstellen für das jeweilige Wort, also jene, welche in allen Trennvarianten vor-

kommen. Eine zweite Ansicht namens *Atomfolgen* zeigt alle Atome der einzelnen Varianten, jeweils durch Punkte getrennt. Die Darstellung *Attributfolgen* (siehe unten) schließlich zeigt die meisten Details und bietet für jede Variante alle Atome und ihre jeweilige Funktionalität in der Wortgrammatik. Wie **Bild 3.2** in einem Ausschnitt aus der Ausgabe des Testsystems zeigt, können diese Erläuterungen bei den höher entwickelten Versionen ReSi und HelSi durchaus umfangreich werden.

```

Attributfolgen (für ReSi und HelSi):
+ wachs=tuben
|...=|...| WortAnfang
`----|--| -> StSbS1 -> StSbS1.
      |--| -> StSbP3-e -> StSbP3-e [=]
      `--| -> SuffE -> StSbP3e.
          ` -> DkSbS2eP2P3e -> WortEnde.

+ wach=stuben
|...=|...| WortAnfang
`----|---| -> StVb -> StVb
      `---| -> StSbP3-e -> StSbP3-e [=]
          ` -> SuffE -> StSbP3e
              ` -> DkSbS2eP2P3e -> WortEnde.

Attributfolgen (für die einfachen Regeln von UrSi):
+ wach=stuben      wach + stub + e + n:
+ S...=S...EE     Stamm + Stamm +Endung + Endung
+ wachs=tuben     wachs + tube + e + n:
+ S...=S...EE     Stamm + Stamm + Endung + Endung

```

Bild 3.2 Attributfolgen-Beispiele für *Wachstuben*

Hierbei ist zu beachten, dass nur zwei Beispiele aus einer großen Zahl sich teilweise bloß in Nuancen unterscheidender Interpretationsvarianten gewählt wurden. Das System muss dennoch alle Varianten genau beachten und darf nur die definitiv sicheren Trennstellen im Endergebnis verwenden.

3.2 Spezielle Methoden

Das Testsystem erlaubt den direkten Vergleich der Resultate aller drei Versionen von SiSiSi und deren intensiven Test.

Das Filtern *unbekannter Wörter* aus der großen Menge analysierter Wörter ist wohl der offensichtlichste Weg, problematische Wörter zu entdecken. Ein Wort wird als *unbekannt* bezeichnet, wenn das System keine gültige Zerlegung und daher auch keine Trennvariante dafür finden kann. Meist sind derartige Wörter nicht richtig geschrieben. Stimmt allerdings die Schreibweise, so hat man eventuell einen Fehler im System, d.h. eine falsche Regel oder ein fehlendes Atom, gefunden. Als dritte Variante kommen geographische Namen und Eigennamen, Abkürzungen oder seltene Wörter, zum Beispiel Fremdwörter, in Frage.

Zusätzlich enthält das Testsystem eine Funktion zum Auffinden aller *mehrdeutigen Wörter*, also jener Wörter, die zumindest zwei unterschiedliche Trennvarianten haben. Unerwünschte Mehrdeutigkeiten, d.h. solche, wo eine der Varianten völlig sinnlos ist, können durch Eintragen des korrekten zusammengesetzten Wortes (ohne Endungen, gewissermaßen als eigener „Wortstamm“) samt Haupttrennstelle in die Atomtabelle beseitigt werden, weil atomare Bestandteile durch den Algorithmus nicht weiter zerlegt werden. Etwaige Nebentrennstellen werden in der Folge wieder automatisch durch die Nachbearbeitung gefunden und müssen daher nicht vermerkt werden.

Werden zwei oder drei Versionen verglichen, so enthält die Ausgabe alle Wörter, für welche die betrachteten Versionen nicht das gleiche Resultat liefern. Der Vergleich von ReSi und HelSi wird in der Hauptsache verwendet, um Wörter zu identifizieren, deren Schreibweise oder Trennung sich durch die Rechtschreibreform geändert hat, während HelSi und UrSi verglichen werden, um den Einfluss der detaillierteren Klassifikation von Atomen und der sehr viel genauer nachgebildeten Grammatikregeln zu studieren.

Die Grammatikregeln erlauben die Verwendung von Trennstrichen in zusammengesetzten Wörtern zur Verbesserung der Lesbarkeit oder zur Sicherstellung der Eindeutigkeit, so wie in *Wach-Stube* oder *Tee-Ei*. Das Testsystem akzeptiert ein Eingabewort mit derartigen Trennstellen („-“) zwischen seinen Bestandteilen oder mit Indikatoren für optionale Trennstellen („~“) und liefert nur jene Trennvarianten, die diese vorgegebenen Trennstellen enthalten. Dadurch können vorgetrennte Wörter als Eingabe verwendet werden, um festzustellen, ob die korrekte Trennung in den von SiSiSi gefundenen Varianten enthalten ist. Mit Hilfe dieser Funktion können Wörter identifiziert werden, die SiSiSi nicht oder falsch trennt, weil z.B. einzelne Atome nicht in der Tabelle enthalten sind.

Das Testsystem schließt auch eine Funktion ein, mit deren Hilfe die Länge der zu analysierenden Wörter auf ein gewisses Intervall beschränkt werden kann. Zum Beispiel kann die Anzahl der Testwörter durch Ignorieren von Wörtern mit weniger als vier Buchstaben eingeschränkt werden, weil es eher unwahrscheinlich ist, dass für diese Wörter falsche Trennstellen entstehen: nach den alten Regeln bleiben sie ungetrennt; nach den neuen Regeln kann nach einem Vokal an erster Stelle in mehrsilbigen Wörtern eine Trennstelle gesetzt werden, die aber wegen ihrer ungünstigen Lage als nicht empfehlenswert markiert wird.

Für die beiden auf detaillierten Grammatikregeln basierenden Versionen, ReSi und HelSi, wurden weitere selektive Tests mit spezialisierten Filterkriterien implementiert. Diese Kriterien wurden speziell gewählt, um die wenigen potentiellen Problemfälle anzusprechen. So ist es etwa möglich, Wörter zu selektieren,



Bild 3.3 Beispiel für Optionen beim Testen von Textdateien mit Spezialkriterien

welche auf Atomen mit wechselnder Funktionalität basieren und daher problematisch sein können; z.B. kann das Atom *ende* als Stamm oder als Endung verwendet werden: *Spiel=en_de* (das Ende eines Spiels), aber auch *Spie-len_de* (spielende Personen).

Zwei Filterkriterien erlauben die Auswahl von Untermengen der Wortbildungs- bzw. Ableitungsregeln, um die korrekte Anwendung dieser Regeln und deren jeweilige spezielle Anwendungsfälle zu untersuchen, wie das Beispiel in **Bild 3.3** neben einigen anderen der vielfältigen Optionen zeigt.

Ein weiteres Kriterium erlaubt die Filterung nach Wörtern, welche atomare Bestandteile spezieller Klassen enthalten. Das ist nützlich, wenn Wörter untersucht werden sollen, die aus bestimmten (zum Beispiel gerade neu in die Klassifizierung aufgenommenen) Atomklassen stammen. Beispiele hierfür wären die Untersuchung von Wortbildungen mit bestimmten Fugenzeichen, Verbleitungen, und so weiter.

Außerdem können auch Wörter aus dem Ergebnis gefiltert werden, welche nur aus einem Atom bestehen, weil diese entweder einsilbig sind und daher ungetrennt bleiben, oder als mehrsilbige Atome basierend auf der Abfolge von Konsonanten und Vokalen getrennt werden. Daher ist für diese Art von Wörtern nicht mit Problemen zu rechnen.

Schließlich kann auch noch nach Wörtern gefiltert werden, die eine bestimmte Zeichenfolge am Anfang, am Ende oder an beliebiger Stelle enthalten. Das hat sich besonders zum Testen der korrekten Klassifizierung neu eingetragener Atome bewährt, weil die Suche nach Kombinationen mit diesen Atomen in großen Testdateien sehr vereinfacht wird.

Ein großer Teil des Testbetriebs mit dem hier vorgestellten System spielte eine wesentliche Rolle bei der fortlaufenden Verbesserung unseres Wortanalyse-systems, zum Beispiel durch die Entdeckung von fehlerhaft oder überhaupt (noch) nicht in die Atomtabelle eingetragenen Wortstämmen (oft Fremdwörtern), die in der Folge

korrekt aufgenommen werden konnten. Der nachstehende Abschnitt zeigt einen kurzen Überblick einiger mit dem aktuellen System durchgeführter Tests.

4 Experimentelle Resultate

Obwohl die Trennung einzelner Wörter durchaus sehr aufschlussreich sein kann, wie das Beispiel in **Bild 4.1** nur ansatzweise demonstriert, ist es natürlich wichtig, möglichst viele Wörter zu testen. Für die Tests wurde daher versucht, große Textdateien zu finden, welche möglichst viele Wortstämme und vorzugsweise alle Wortarten der deutschen Sprache abdecken. Leider waren die meisten Wörterbücher für die Testläufe nicht in geeigneter Form verfügbar oder enthielten nicht die gewünschte Vielfalt an Wortformen, daher mussten andere geeignete Dateien gefunden werden. Es wurde eine Testdatei mit mehr als 209.000 verschiedenen Wortformen² verwendet, um nach Wörtern zu suchen, die von HeSi nicht erkannt werden. Anfangs wurden etwa 12,5% als unbekannt gemeldet.

² verfügbar unter <http://www.sibiller.de/anagramme/>

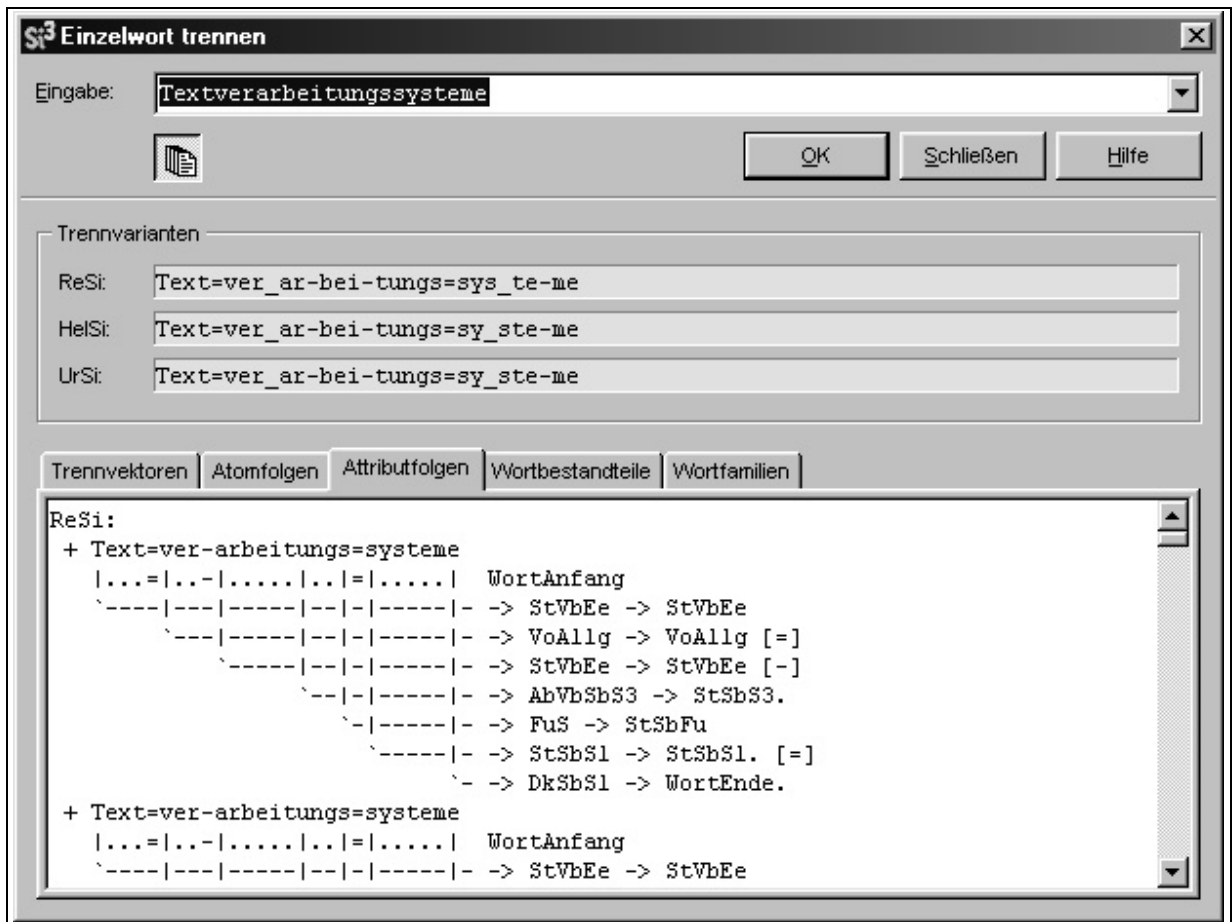


Bild 4.1 Trennung eines Einzelwortes im Testsystem mit Anzeige der Attributfolgenliste

Diese Wörter wurden genauer überprüft und die Atomtabelle entsprechend aktualisiert. Hierbei ist zu beachten, dass die Testdatei eine große Menge von flektierten Wortformen, Ableitungen und zusammengesetzten Wörtern enthält. Dadurch ergab sich beim Einfügen eines einzelnen Atoms bereits die Elimination einer großen Anzahl von Problemfällen. Unter den unbekanntem Wörtern befanden sich biographische und geographische Namen sowie daraus abgeleitete Wortformen. Außerdem ist die Anzahl als unbekannt gemeldeter Wörter etwas höher, weil es in der Testdatei falsch geschriebene oder abgekürzte Wörter gibt, die das System korrekterweise als nicht korrekt gebildet erkennt.

Mit Hilfe einer Liste aus [5]³, die Wörter und Phrasen enthält, deren Schreibweise sich durch die Rechtschreibreform geändert hat, wurden ReSi und HelSi im Vergleich getestet. Aus dieser Liste wurden zwei Testdateien von jeweils rund 2.000 Wörtern erzeugt, welche Wörter in alter bzw. neuer Schreibweise enthalten. ReSi und HelSi analysierten jeweils die Wörter mit neuer bzw. alter Schreibung korrekt, während die

Wörter der jeweils anderen Schreibweise als fehlerhaft markiert wurden. Auch die geänderten Trennungsvorschriften für Buchstabenfolgen wie *st* und *ck* wurden jeweils korrekt angewandt.

In **Bild 4.2** ist zu sehen, dass nur ReSi den nach neuen Regeln geschriebenen *Delfin* erkennt. Alle drei Versionen erkennen und trennen hingegen den klassisch geschriebenen *Delphin* problemlos.



Bild 4.2 Beispiel für die Analyse von *Delfin*

³ auch verfügbar unter <http://www.duden.de/>

Des weiteren wird unter ReSi auch die neue Regel zur Silbentrennung angewendet, wonach bestimmte Wörter entweder wie bisher entsprechend der Trennung in ihrer Herkunftssprache oder auch nach den im Deutschen üblichen Konsonanten-Vokal-Folgen getrennt werden dürfen. Wie die Trennvarianten *Ma-gne-ti-sie-rung* (alt) und *Mag-ne-ti-sie-rung* (neu) durch unterschiedliche Trennzeichen in einem Wort eingetragen werden, zeigt **Bild 4.3**. Alle Trennstellen, die in beiden Versionen an der gleichen Stelle vorkommen, werden auf die gewohnte Weise angezeigt.

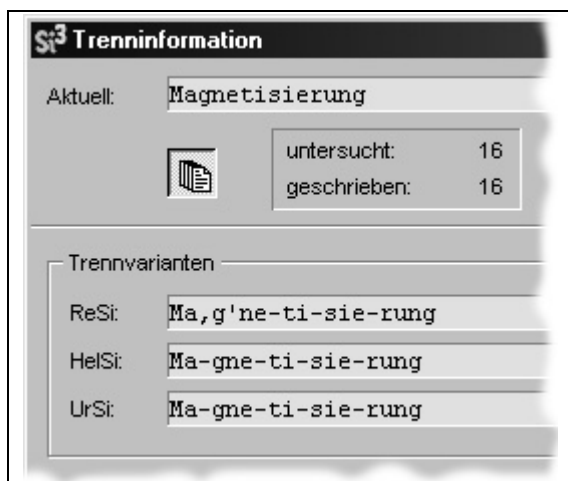


Bild 4.3 Trennvarianten von *Magnetisierung*

Es ist geplant, mit Hilfe mehrerer anderer Testdateien, welche unter anderem durch Exportieren bestimmter Wortklassen aus dem Datenbestand des morphologischen Analysators *Morphy* [4] gewonnen wurden, weitere Tests und Verbesserungen durchzuführen. Wie schon bisher ist eines der großen Probleme leider die Beschaffung qualitativ hochwertiger Testdaten in brauchbarer Form.

5 Ausblick

Andere auf Wortanalyse basierende Anwendungen können leicht in das Testsystem integriert und in der Folge ebenfalls wie beschrieben umfassend getestet werden. Ein Beispiel ist die *sinnentsprechende Volltextsuche*, welche den Algorithmus zur Wortanalyse verwendet, um eine Menge von Atomen zu erhalten, welche die Bedeutung eines gegebenen Eingabewortes bestimmen [2], siehe auch [3]. Eine Funktion, welche die Menge sinngebender Bestandteile liefert, wurde für alle drei Versionen von SiSiSi bereits in das Testsystem integriert.

Es besteht die Absicht, ähnliche Erweiterungen auch zur Rechtschreibkontrolle und zur Prüfung oder (Wieder-)Herstellung der korrekten Großschreibung zu implementieren. Ein großer Vorteil liegt auch hier wieder

in der Tatsache, dass SiSiSi nur eine relativ kleine Atomtabelle samt Grammatikregeln braucht, um eine große Menge von Wörtern abzudecken.

6 Zusammenfassung

Die vorgestellte Testumgebung wird verwendet, um die sichere, sinnensprechende Silbentrennung und das ihr zugrundeliegende Wortanalyzesystem zu testen. Sie ermöglicht den intensiven Test und den Vergleich dreier unterschiedlicher Verfahrensvarianten, sowohl in Verbindung mit den alten als auch mit den neuen Rechtschreibregeln. Im Rahmen des intensiven Testens wurde eine Reihe spezieller Filterkriterien vorgestellt, durch deren Anwendung insbesondere beim Testen von großen Textdateien die wenigen potentiellen Problemfälle gezielt herausgefiltert werden können. Mit Hilfe dieser Testumgebung wurden bereits umfangreiche Tests durchgeführt, die entscheidend zur Verbesserung des Wortanalyzesystems beitragen, indem beispielsweise fehlende oder nicht korrekt klassifizierte Stämme entdeckt und diese Mängel in der Folge behoben werden konnten.

7 Literatur

- [1] Barth, W., Nirschl H.: Sichere sinnensprechende Silbentrennung für die deutsche Sprache. *Angewandte Informatik* 4, S. 152–159, 1985
- [2] Barth, W.: Volltextsuche mit sinnensprechender Wortzerlegung. *Wirtschaftsinformatik*, 32. Jahrgang, Heft 5, S. 467–471, 1990
- [3] Kodydek, G.: Ein Wortanalyzesystem für Silbentrennung, Volltextsuche und Rechtschreibprüfung unter Berücksichtigung der Rechtschreibreform. In: *Tagungsband Konvens'2000*, Ilmenau, Deutschland, 2000
- [4] Lezius, W., Rapp, R., Wettler, M.: A Freely Available Morphological Analyzer, Disambiguator, and Context Sensitive Lemmatizer for German. In: *Proceedings of the COLING-ACL 1998*, Montreal, Canada, 1998
- [5] Sitta, H., Gallmann, P.: *Duden, Informationen zur neuen deutschen Rechtschreibung*, Hrsg. Dudenredaktion. Dudenverlag, Mannheim, 1996
- [6] Steiner, H.: *Automatische Silbentrennung durch Wortbildungsanalyse*. Dissertation, Institut für Computergraphik, Technische Universität Wien, 1995
- [7] Steiner, H., Barth, W.: Sichere sinnensprechende Silbentrennung mit Berücksichtigung der deutschen Wortbildungsgrammatik. In: *Tagungsband Konvens'94*, Hrsg. H. Trost, Wien, S. 330–340, 1994