

MORPHY für Windows
Ein Tool zur Deutschen Morphologie
- Benutzerhandbuch -

Wolfgang Lezius, IMS, Universität Stuttgart

10. März 1999

Inhaltsverzeichnis

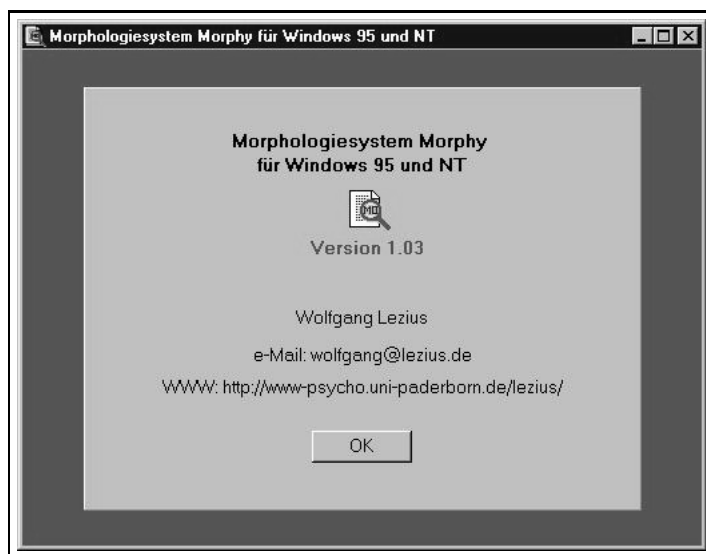
1	Programmstart und Menüpunkte	2
1.1	Programmstart	2
1.2	Das Hauptmenü	2
2	Morphologische Analyse	5
2.1	Analyse einer Wortform	5
2.2	Analyse einer Datei	6
2.3	Optionen	6
3	Tagging	9
3.1	Tagging eines Satzes	9
3.2	Tagging einer Datei	10
3.3	Optionen	11
4	Generierung	12
4.1	Lookup	12
4.2	Generierung aus einer Datei	14
5	Lexikon	16
5.1	Lexikonpflege	16
5.2	Statistik über das Lexikon	18
5.3	Export des Morphy-Lexikons	18

Kapitel 1

Programmstart und Menüpunkte

1.1 Programmstart

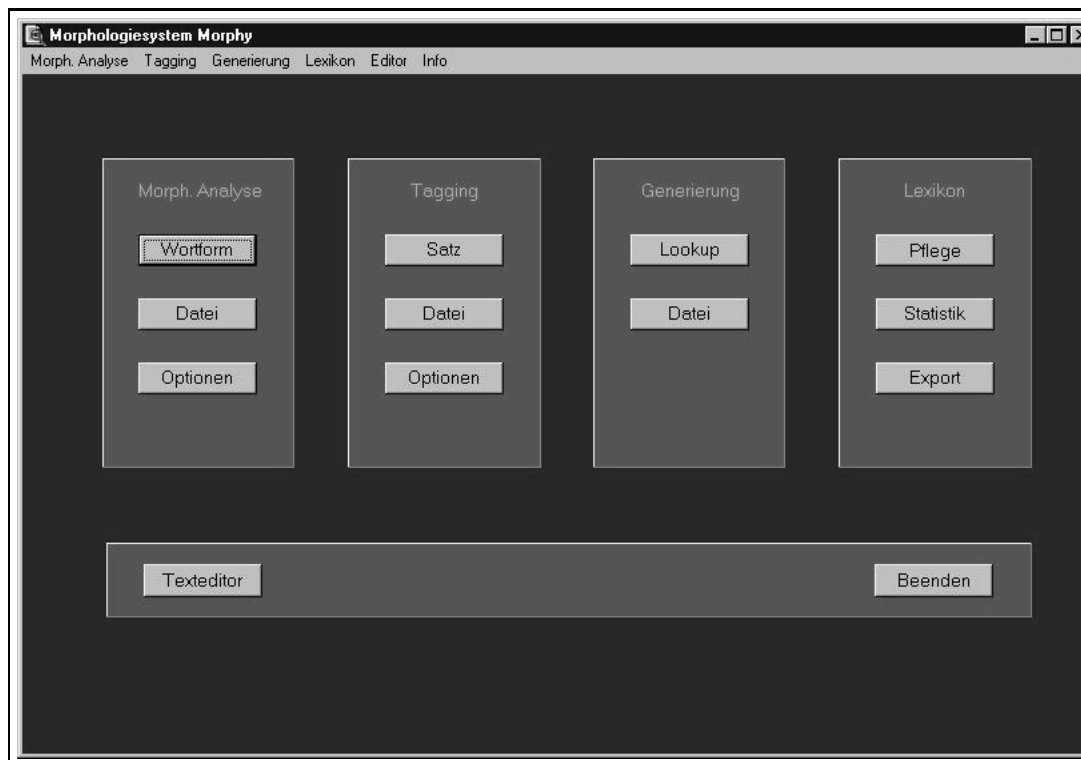
Sie starten Morphy durch Anklicken des Symbols in der Programmgruppe oder durch Aufruf von `morphy.exe`. Nun erscheint der Begrüßungsbildschirm, der unter anderem auf die aktuelle Programmversion hinweist:



Bestätigen Sie durch Anklicken des OK-Buttons oder Betätigen der Eingabetaste. Bevor das Hauptmenü erscheint, werden die Lexikondaten in den Hauptspeicher geladen. Dieser Vorgang dauert zwei bis drei Sekunden. Beim allerersten Programmstart wird die interne Speicheroptimierung initialisiert, was zusätzlich etwa fünf Sekunden in Anspruch nimmt.

1.2 Das Hauptmenü

Das Hauptmenü ist in vier Bereiche unterteilt:

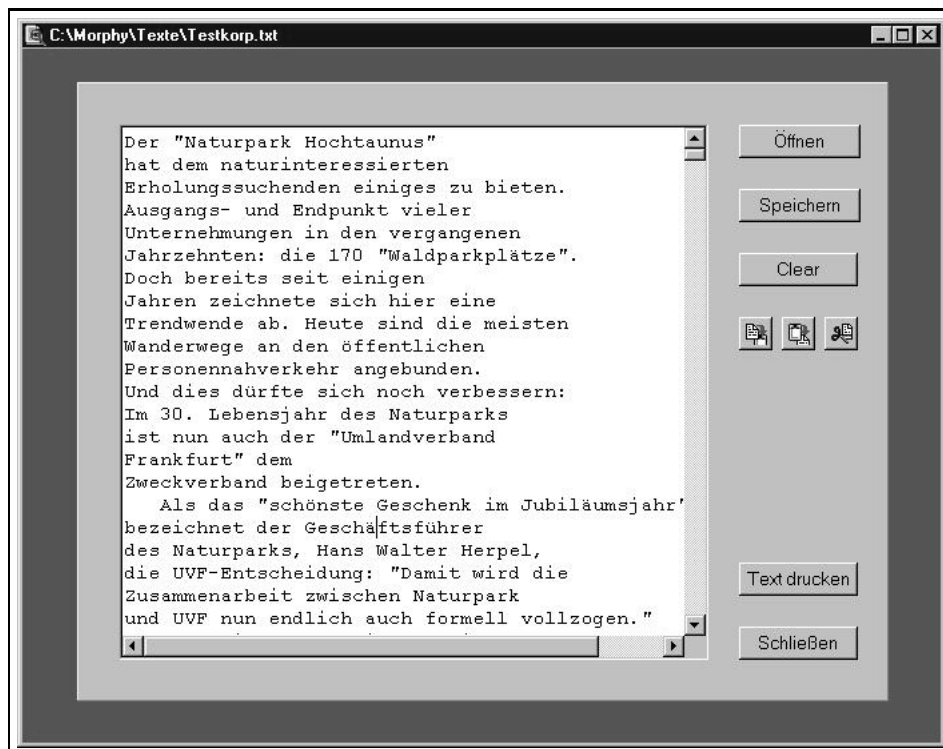


- Morphologische Analyse
- Tagging
- Morphologische Generierung
- Lexikon

Alle Punkte sind sowohl über das Pull-Down-Menü der Fensterleiste sowie direkt über Buttons erreichbar. In den folgenden Kapiteln werden die vier Bereiche ausführlich beschrieben.

Sie beenden das Programm entweder über den *Beenden*-Button, durch Anklicken der linken oberen Ecke oder durch den Punkt *Beenden* im Pull-Down-Menü der *Morphologischen Analyse*.

Da es unter Windows etwas mühsam ist, ein Editorfenster neben einer Anwendung zu benutzen, ein Editor zum Anschauen von Teilergebnissen von Morphy aber sehr hilfreich ist, ist zusätzlich ein kleiner Texteditor integriert. Wenn Sie den Texteditor über den *Texteditor*-Button unten links im Hauptmenü aufrufen, erscheint das Editor-Fenster:



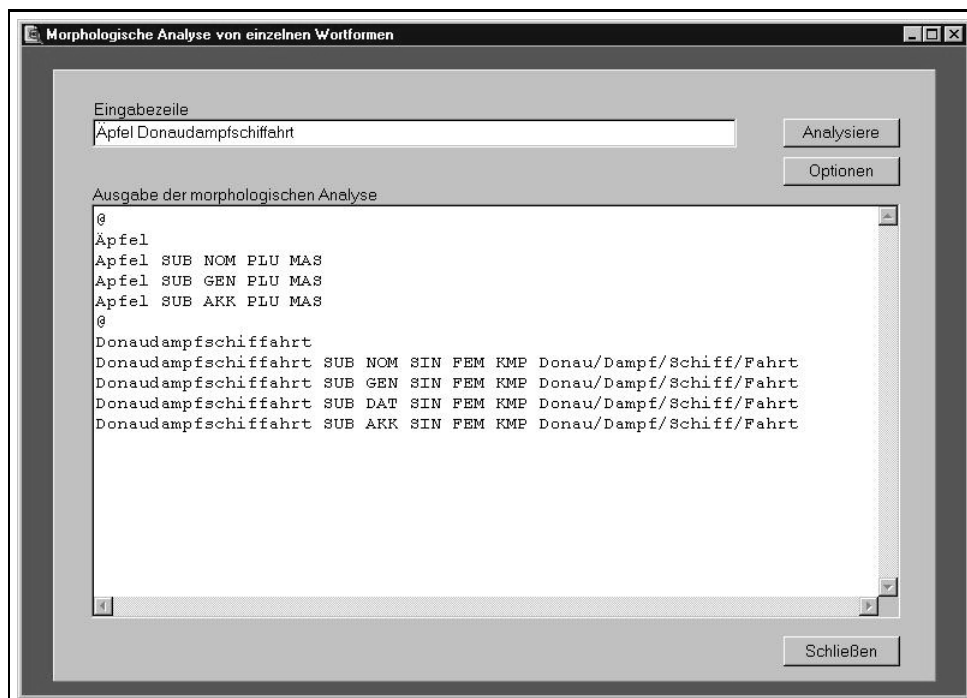
Neben den obligatorischen Funktionen wie dem Öffnen und Speichern einer Datei, kann der Inhalt des Editors auch jederzeit auf dem aktuell gültigen Standarddrucker ausgedruckt werden. Die Toolbar erlaubt das Kopieren (auch über die Tastenkombination CTRL-C) oder Ausschneiden (CTRL-X) eines markierten Textausschnitts in die Zwischenablage sowie das Einfügen der Zwischenablage (CTRL-V) an der aktuellen Cursorposition.

Kapitel 2

Morphologische Analyse

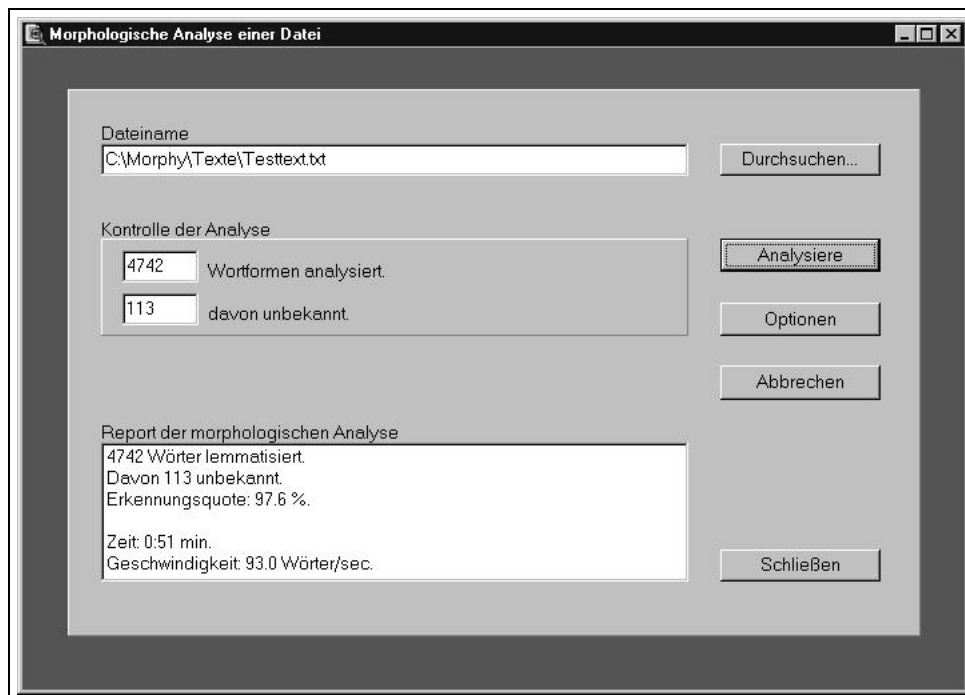
Die Hauptanwendung der Analyse-Komponente besteht in der morphologischen Analyse einer vorliegenden Textdatei. Doch um sich ein Bild von der Leistungsfähigkeit der Morphologie zu machen, ist eine direkte Analyse von Hand eingetippter Wortformen enorm hilfreich. Aus diesem Grunde wurde der Punkt *Analyse einer Wortform* implementiert. Alle Optionen (vgl. Punkt 2.3) können sowohl vom Hauptmenü als auch von den beiden Analyse-Menüpunkten aus eingestellt werden.

2.1 Analyse einer Wortform



In der Eingabezeile können Sie beliebig viele Wortformen eintippen, die dann nach Betätigen des *Analyse*-Buttons Wort für Wort analysiert werden. Das Ausgabeformat können Sie dabei in den Optionen verändern.

2.2 Analyse einer Datei



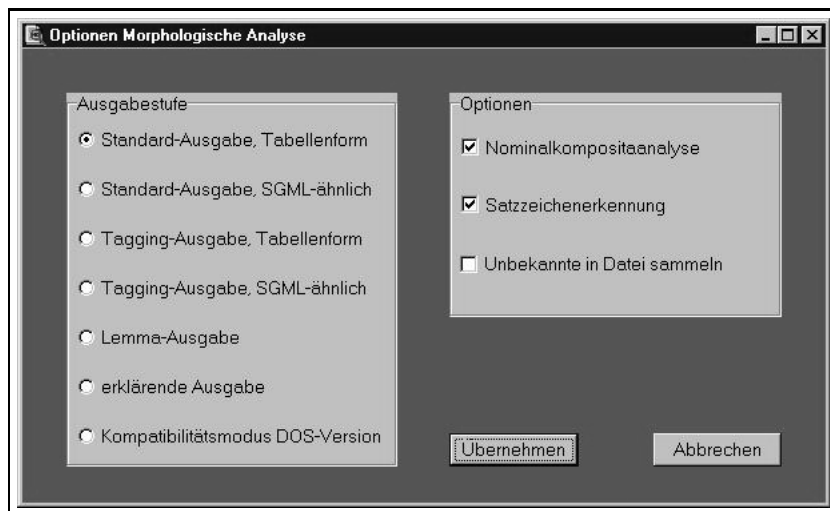
Zunächst muß die zu analysierende Datei benannt werden. Klicken Sie dazu auf den *Durchsuchen*-Button und bestimmen Sie die Datei im nun folgenden Windows-Dateidialog.

Durch Drücken des *Analysiere*-Buttons wird der Vorgang gestartet. Zur Kontrolle werden die Anzahl analysierter Wortformen und nicht erkannter Wortformen ausgegeben. Der Vorgang läßt sich jederzeit durch Anklicken von *Abbrechen* stoppen. Nach Beendigung der Analyse werden die Ergebnisse im Report-Bereich zusammengefaßt (z.B. Geschwindigkeit der Analyse).

Die Resultate der Analyse werden in der Datei `dateiname.lem` abgelegt, im selben Verzeichnis wie die Ausgangsdatei `dateiname.endung`.

2.3 Optionen

Das Optionsfenster ist in zwei Bereiche unterteilt: Im einen Bereich wird die Ausgabestufe eingestellt, d.h. die Tiefe der morphologischen Analyse. Im zweiten Bereich können einige Feinheiten der Analyse parametrisiert werden.



Die Standardausgabe umfasst alle verfügbaren morphosyntaktischen Informationen. Die Formatierung ist standardisiert, um eine Weiterverarbeitung zu ermöglichen. Die Taggingausgabe beschränkt sich auf die wesentlichen Attribute, die Lemmaausgabe gibt nur die Lemmakandidaten an. Wenn eine Lesart nur am Satzanfang gültig ist (z.B. die Wortform *Bist*), ist der jeweilige Lemmakandidat mit einem Stern markiert (also z.B. **sein*).

Bei den übrigen Optionen können Sie die Nominalkompositaanalyse ein- und ausschalten, Satzzeichen ignorieren und bei Bedarf alle Wortformen, die Morphy nicht analysieren kann, in der Datei `dateiname.unb` sammeln (nur bei der Analyse einer Datei).

Die folgenden beiden Tabellen zeigen alle verfügbaren Ausgabestufen und veranschaulichen diese anhand der Analyse der Wortform *Hauses*. Eine detaillierte Beschreibung finden Sie in einem Paper auf der Morphy-Website.

Analysestufe	Tiefe der Beschreibung	Formatierung
Standardausgabe, Tabellenform	alle verfügbaren Attribute	Attribute nebeneinander, tabellenartig
Standardausgabe, SGML-ähnlich	alle verfügbaren Attribute	Attribut-Werte-Paare, SGML-Formatierung
Taggingausgabe, Tabellenform	wesentliche Attribute	Attribute nebeneinander, tabellenartig
Taggingausgabe, SGML-ähnlich	wesentliche Attribute	Attribut-Werte-Paare, SGML-Formatierung
Lemmaausgabe	Lemma(ta)	nebeneinander
erklärende Ausgabe	alle verfügbaren Attribute	im Klartext, sehr ausführlich
DOS-Kompatibilitäts-Modus	alle verfügbaren Attribute	ähnlich Standardausgabe Reihenfolge variiert

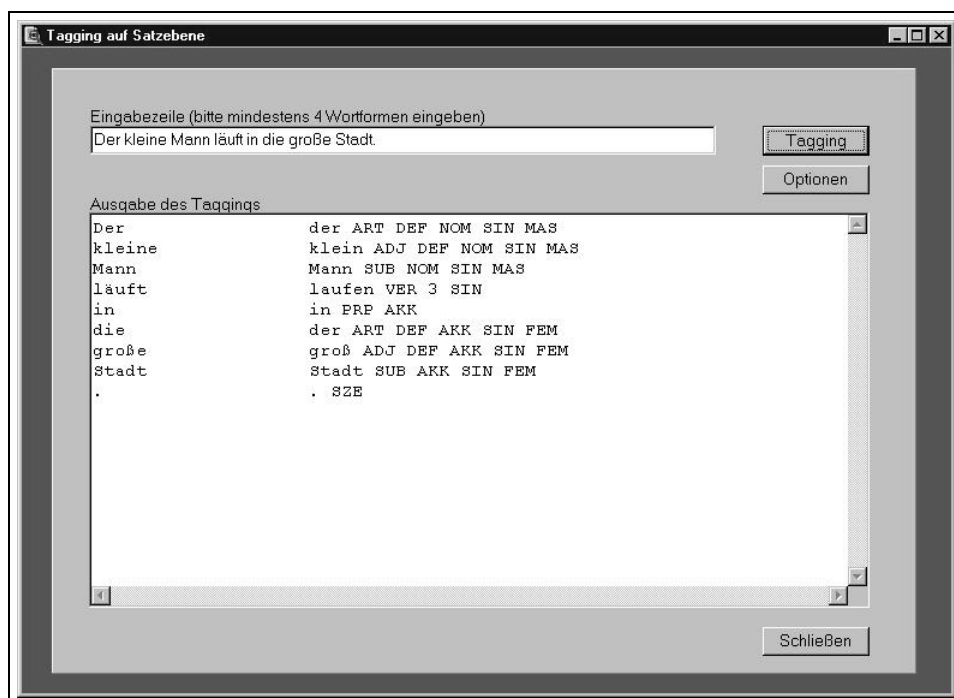
Analysestufe	Beispiel
Standardausgabe, Tabellenform	Haus Haus SUB GEN SIN NEU
Standardausgabe, SGML-ähnlich	<form>Houses</form> <lemma wkl=SUB kas=GEN num=SIN gen=NEU>Haus</lemma>
Taggingausgabe, Tabellenform	Houses Haus SUB GEN SIN NEU
Taggingausgabe, SGML-ähnlich	<form>Houses</form> <lemma wkl=SUB kas=GEN num=SIN gen=NEU>Haus</lemma>
Lemmaausgabe	Houses - Haus
erklärende Ausgabe	Houses Substantivform von Haus Genitiv Singular (maskulinum)
DOS-Kompatibi- litäts-Modus	Houses SUB Haus GEN NEU SIN

Kapitel 3

Tagging

Analog zur morphologischen Analyse ist auch hier ein spezieller Modus eingebaut, in dem Sie einen Satz oder Satzbestandteil eintippen und das Tagging-Resultat direkt einsehen können. Die Lemmatisierung des Textes geschieht während des Taggings und braucht deshalb nicht separat aktiviert zu werden.

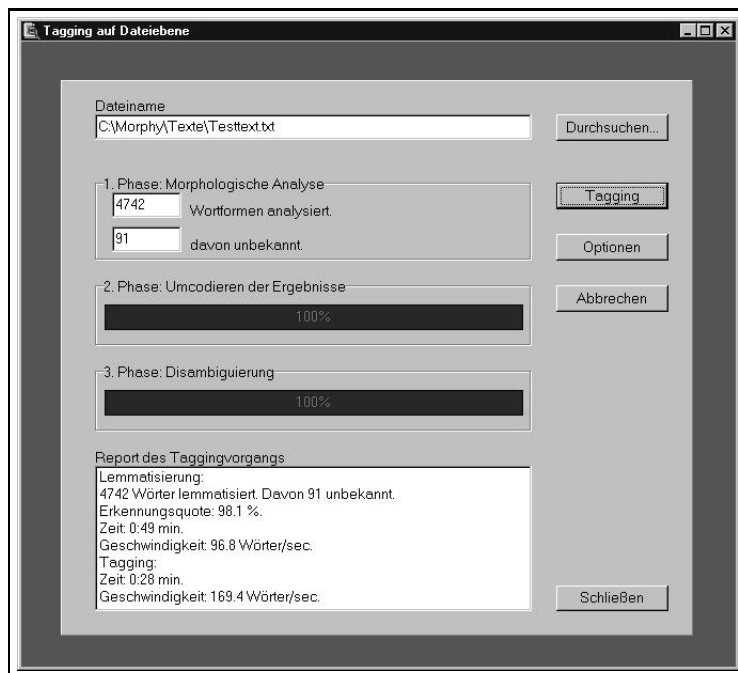
3.1 Tagging eines Satzes



Geben Sie den Satz oder den Satzbestandteil in der Eingabezeile ein. Die Eingabe muß mindestens vier Wortformen (inkl. Satzzeichen) umfassen. Durch Betätigen des *Tagging*-Buttons wird der Tagging-Prozess angestoßen. Die Ergebnisse können Sie im Ausgabebereich einsehen.

3.2 Tagging einer Datei

Das Taggingfenster ist analog zur Morphologie aufgebaut:



Geben Sie zunächst wieder durch Anklicken des *Durchsuchen*-Buttons die zu taggende Datei an. Stellen Sie ggf. die Optionen neu ein (vgl. Punkt 3) und betätigen Sie anschließend den *Tagging*-Button. Der Tagging-Vorgang wird nun gestartet.

Im ersten Schritt wird die morphologische Analyse durchgeführt. Es handelt sich um eine Vorstufe des Taggings, die in der Regel eine sehr große Ergebnisdatei produziert. Sie wird deshalb nach Beendigung des Taggings wieder gelöscht. Falls jedoch eine entsprechende Option gewählt wurde (vgl. Punkt 3.3), werden die Ergebnisse der morphologischen Analyse in der Datei `dateiname.1et` zwischengespeichert. Falls Sie dieselbe Datei später in einem anderen Format oder mit einem anderen TagSet taggen wollen, sparen Sie hierdurch die erneute morphologische Analyse.

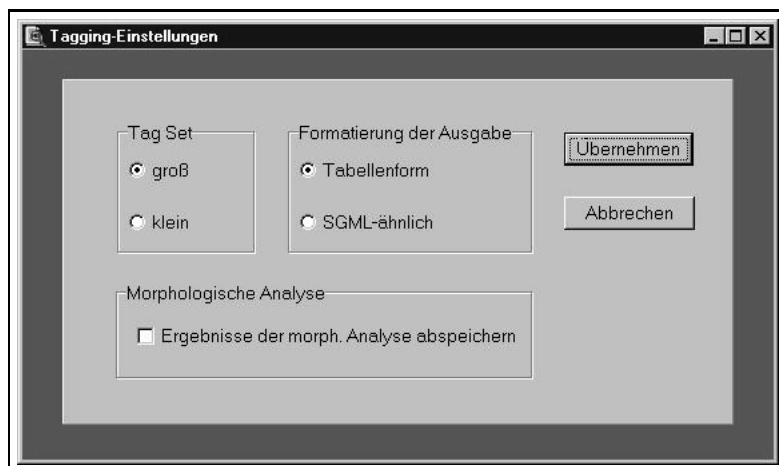
In einem zweiten Schritt wird die Analyse in das Tagging-Format umcodiert. Dieser Vorgang ist architekturbedingt und spielt für die Bedienung keine Rolle.

Im dritten und letzten Schritt erfolgt nun die Disambiguierung und damit das eigentliche Tagging.

Jeden dieser drei Schritte können Sie jederzeit durch Betätigen des *Abbrechen*-Buttons stoppen. Die Ergebnisse des Taggings (Geschwindigkeit usw.) werden im Report-Bereich zusammengefaßt. Die Tagging-Resultate werden in der Datei `dateiname.sma` bzw. `dateiname.big` gespeichert, abhängig vom TagSet (*small* oder *large*).

3.3 Optionen

Für das Tagging können das verwendete TagSet und die Formatierung der Ausgabe parametrisiert werden. Das große TagSet umfaßt die wesentlichen morphosyntaktischen Attribute wie Kasus, Numerus und Genus, das kleine TagSet beschränkt sich auf die Angabe der Wortklasse.



Um einen Eindruck zu bekommen, welche Tiefe die TagSets bieten, ist hier ein Beispiel für jede Ausgabe aufgeführt. Ausführlich beschrieben werden die TagSets in einigen Papers, die auf der Morphy-Website zu finden sind.

TagSet	Formatierung	Beispiel
klein	Tabellenform	der - der ART DEF kleine - klein ADJ Mann - Mann SUB
klein	SGML-ähnlich	<wort>der</wort> <lemma tag=ART DEF>der</lemma> <wort>kleine</wort> <lemma tag=ADJ>klein</lemma> <wort>Mann</wort> <lemma tag=SUB>Mann</lemma>
groß	Tabellenform	der - der ART DEF NOM SIN MAS kleine - klein ADJ DEF NOM SIN MAS Mann - Mann SUB NOM SIN MAS
groß	SGML-ähnlich	<wort>der</wort> <lemma wkl=ART typ=DEF kas=NOM num=SIN gen=MAS>der</lemma> <wort>kleine</wort> <lemma wkl=ADJ art=DEF kas=NOM num=SIN gen=MAS>klein</lemma> <wort>Mann</wort> <lemma wkl=SUB kas=NOM num=SIN gen=MAS>Mann</lemma>

Neben der Taggingart können Sie noch einstellen, ob die Ergebnisse der morphologischen Analyse, die für das Tagging notwendig sind, in der Datei `dateiname.let` zwischengespeichert werden (vgl. Punkt 3.2).

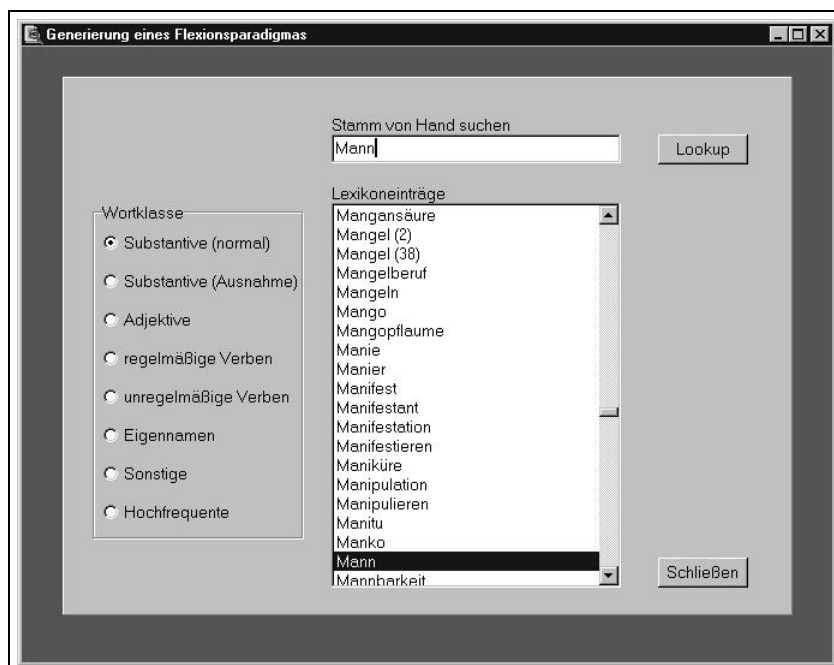
Kapitel 4

Generierung

Die Generierungs-Funktion wird in Morphy auf zwei Arten benutzt. Zum einen erlaubt sie ein Browsen durch die einzelnen Lexika. Sie können einen Stamm auswählen und sein Flexionsparadigma einsehen. Desweiteren kann das Flexionsparadigma formatiert in eine Datei gelenkt werden.

4.1 Lookup

Nach der Anwahl der Generierung können Sie zunächst ein Unterlexikon, dann einen Stamm aus einer Stammliste auswählen:



Im Feld *Stamm von Hand suchen* können Sie die Stammliste anhand von Anfangsbuchstaben durchsuchen. Tippen Sie z.B. *geh* bei unregelmäßigen Verben ein, springen Sie in der Liste zu *gehen*. Dies erleichtert das Auffinden von Stämmen. Klicken Sie nun per

Doppelklick auf den gewünschten Stamm oder auf den *Lookup*-button, so erscheint das Flexionsparadigma des Stamms. Die folgenden drei Abbildungen zeigen die Paradigmen des Verbs *schlagen*, des Adjektivs *edel* und des Substantivs *Mann*.

Konjugation eines Verbs

Konjugation des Verbs: schlagen

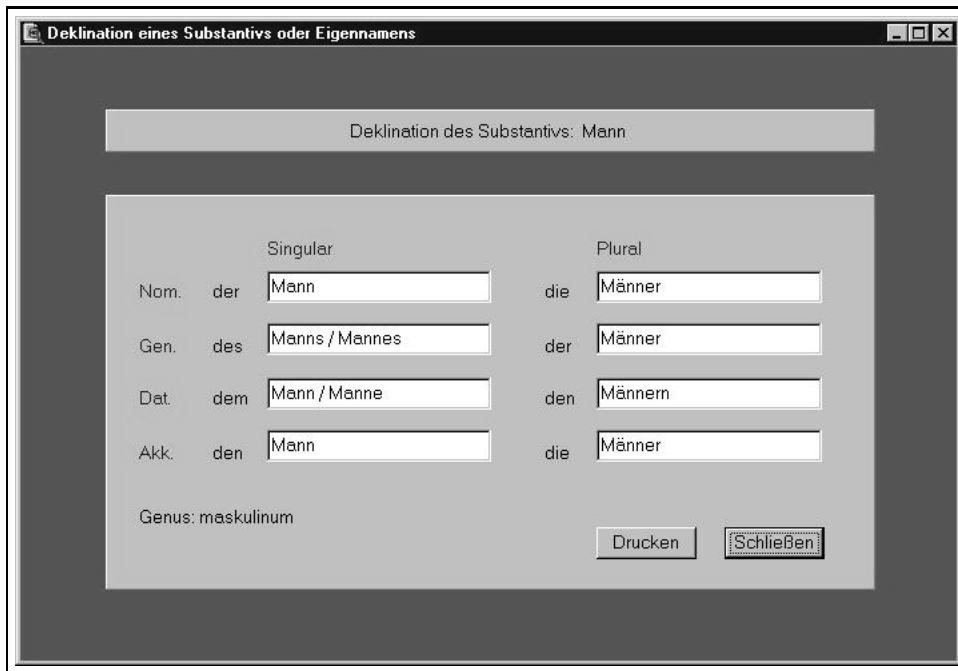
<p>Präsens</p> <p>ich schlage du schlägst er schlägt wir schlagen ihr schlagt sie schlagen</p>	<p>Konjunktiv I</p> <p>ich schlage du schlägest er schlage wir schlagen ihr schlaget sie schlagen</p>	<p>Partizipien</p> <p>schlagend habe/bin geschlagen</p>
<p>Imperfekt</p> <p>ich schlug du schlugst er schlug wir schlugen ihr schlugt sie schlugen</p>	<p>Konjunktiv II</p> <p>ich schlüge du schlügest er schlüge wir schlügen ihr schlüget sie schlügen</p>	<p>Imperative</p> <p>schlage! schlagt!</p>
		<p>Präfixe</p> <p>be er ver zer</p>
		<p>Zusätze</p> <p>ab an auf aus breit durch ein entzwei fehl</p>

Drucken Schließen

Deklination eines Adjektivs

	Positiv edel	Komparativ edler	Superlativ am edelsten
	Positiv maskulinum	femininum	neutrum
ohne Artikel	edler edlen edlem edlen edle edler edlen edle	edle edler edler edle edle edler edlen edle	edles edlen edlem edles edle edler edlen edle
best. Artikel	der edle des edlen dem edlen den edlen die edlen der edlen den edlen die edlen	die edle der edlen der edlen die edle die edlen der edlen den edlen die edlen	das edle des edlen dem edlen das edle die edlen der edlen den edlen die edlen
unbest. Artikel	ein edler eines edlen einem edlen einen edlen keine edlen keiner edlen keinen edlen keine edlen	eine edle einer edlen einer edlen eine edle keine edlen keiner edlen keinen edlen keine edlen	ein edles eines edlen einem edlen ein edles keine edlen keiner edlen keinen edlen keine edlen

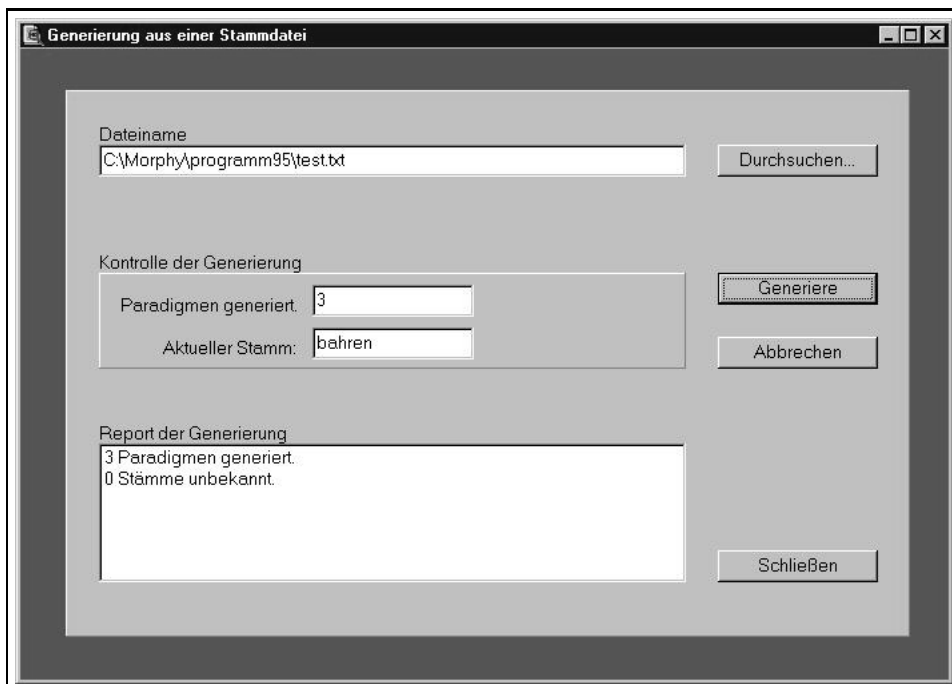
Positiv Komparativ Superlativ Drucken Schließen



Wenn Sie zwischendurch das Lexikon pflegen, verläuft der Aufbau der Wortlisten für Adjektive und Substantive beim ersten Aufruf etwas langwierig. Hier wird ein Cache-Mechanismus initialisiert, der die Ausführung von diesem Zeitpunkt an enorm beschleunigt.

4.2 Generierung aus einer Datei

Geben Sie zunächst wieder die Datei, aus der generiert werden soll, durch Anklicken von *Durchsuchen* an.



Die Datei muß die gewünschten Stämme zeilenweise **untereinander** enthalten. Starten Sie den Generierungs-Prozeß durch Anklicken von *Generierung*. Zur Kontrolle des Vorgangs wird der jeweils aktuelle Stam ausgegeben. Die Resultate werden im Report-Fenster zusammengefaßt. Optionen sind nicht einstellbar.

Kapitel 5

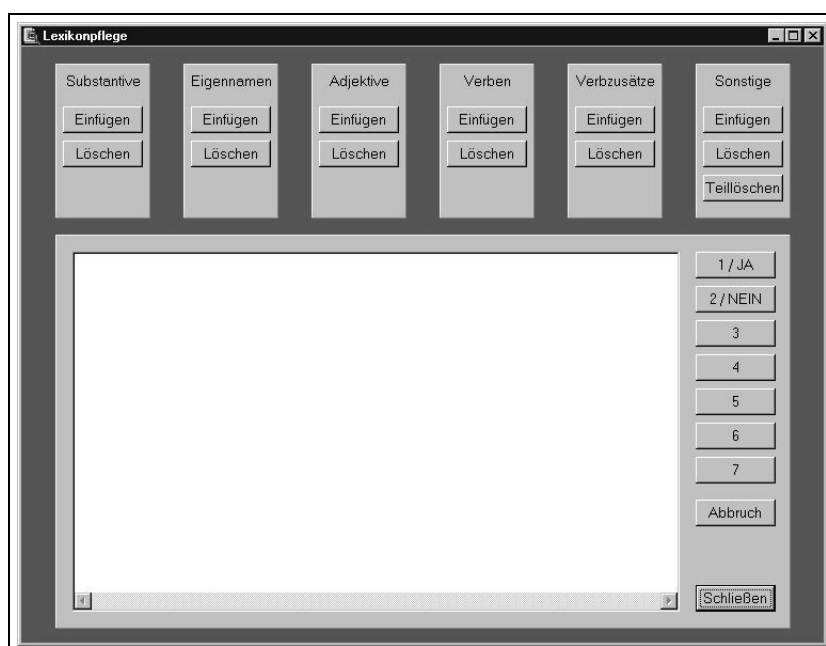
Lexikon

Der Lexikonbereich besteht im wesentlichen aus der Lexikonpflege und einer Übersicht über das Lexikon. Da Morphy nicht unter Unix lauffähig ist, viele Anwender aber nur unter Unix arbeiten, ist zudem eine Exportfunktion implementiert.

5.1 Lexikonpflege

Die Pflege des Morphy-Lexikons ist zwar auch für Laien problemlos durchführbar. Umfangreichere Pflegevorgänge sind aber vermutlich nicht nötig. Das Morphy-Lexikon dürfte den Grundwortschatz der Deutschen Sprache abdecken, was für einfache Anwendungen völlig ausreichen sollte. Wollen Sie dennoch das Lexikon pflegen, finden Sie hier einige grundlegende Hinweise zum Umgang mit der Pflegekomponente:

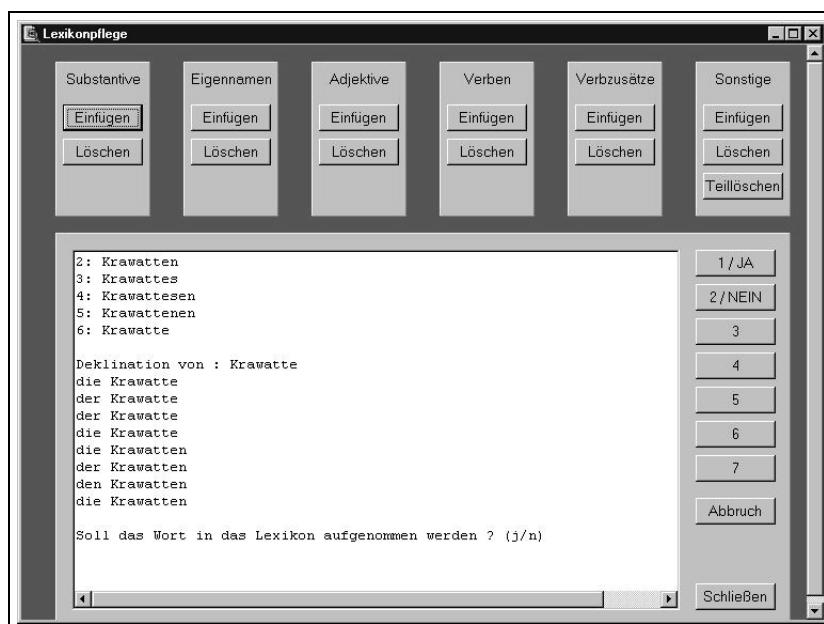
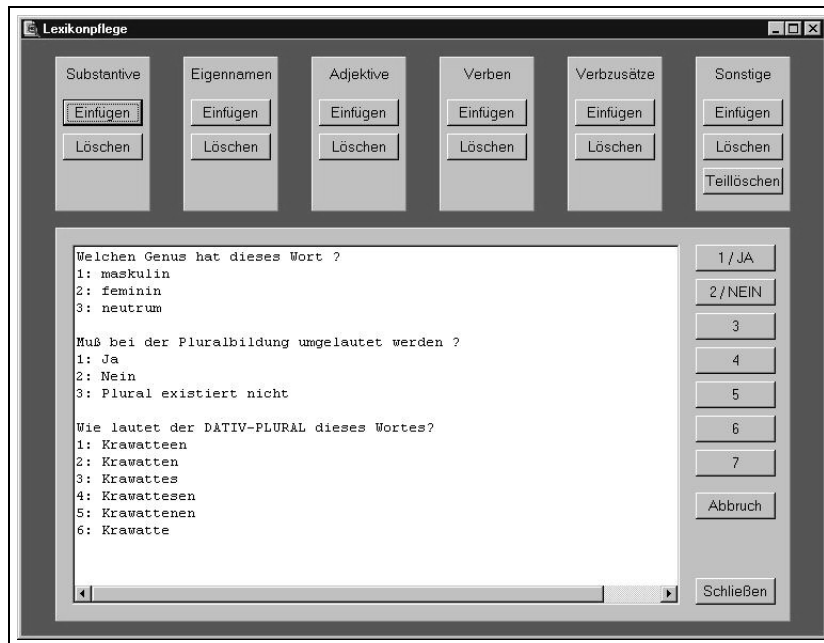
Wenn Sie im Hauptmenü den Punkt *Lexikonpflege* anwählen, erscheint das Lexikonpflege-Fenster, das in zwei Bereiche unterteilt ist:



Im oberen Bereich wählen Sie den gewünschten Aktionstyp aus, z.B. Einfügen oder Löschen eines Substantivs. Im unteren Bereich findet der eigentliche Dialog statt, der zum Eintragen oder Löschen eines Stamms führt.

Einfügen eines Stamms

Das folgende Beispiel soll die Vorgehensweise erläutern. Klicken Sie auf *Substantiv einfügen*. Sie werden zunächst nach dem Stamm des Substantivs gefragt. Nun folgt der Eintrags-Dialog, im Laufe dessen Sie Fragen zum Flexionsmuster des Substantivs beantworten müssen. Die folgenden beiden Abbildung zeigen einen solchen Dialog für das Substantiv *Krawatte*:



Löschen eines Stamms

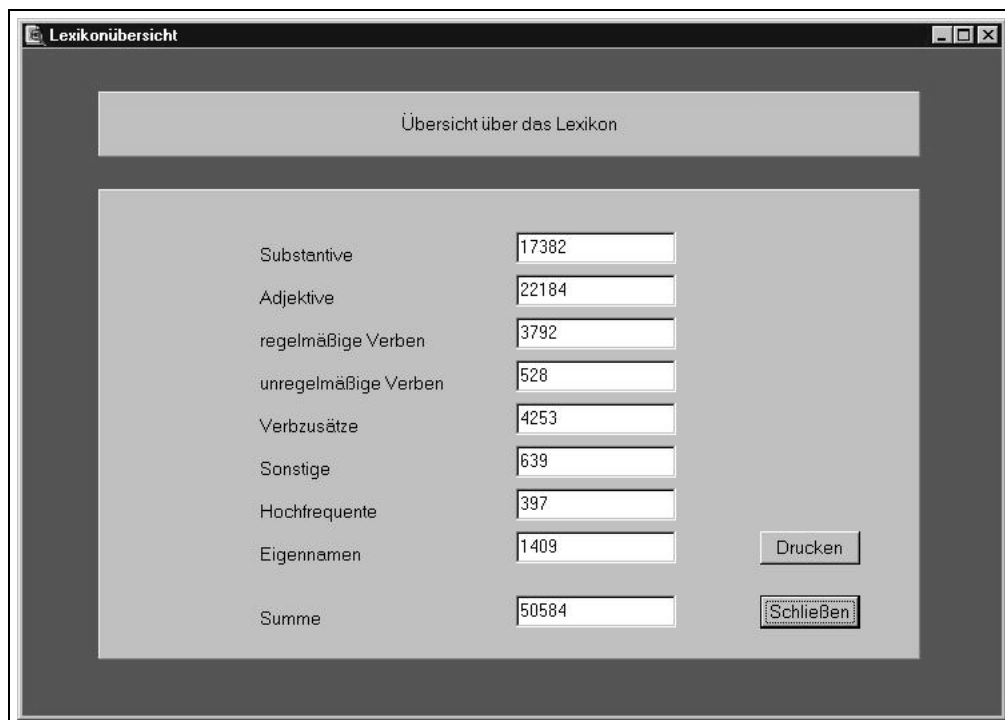
Klicken Sie die Aktion an. Geben Sie den zu löschenden Stamms ein - fertig.

Einfügen oder Löschen eines Verbzusatzes

Ein sogenannter *Verbzusatz*, häufig auch als abtrennbares Verbpräfix bezeichnet, ist mit einem Verb verbunden (z.B. *ab* und *spielen* zu *abspielen*). Die Verbindung wird im Hauptsatz aufgehoben (*er spielt den Ball ab*), bleibt im Nebensatz dagegen erhalten (*wenn er den Ball abspielt.*). Wenn ein solcher Verbzusatz eingefügt werden soll, ist deshalb immer auch die Angabe des Verbs erforderlich.

5.2 Statistik über das Lexikon

Durch Anklicken von *Statistik* im Hauptmenü erhalten Sie eine detaillierte Übersicht über die Wortklassen-Lexika:



The screenshot shows a window titled 'Lexikonübersicht' with a header 'Übersicht über das Lexikon'. Below the header is a table with two columns: word classes and their counts. The counts are displayed in input fields. To the right of the table are two buttons: 'Drucken' and 'Schließen'.

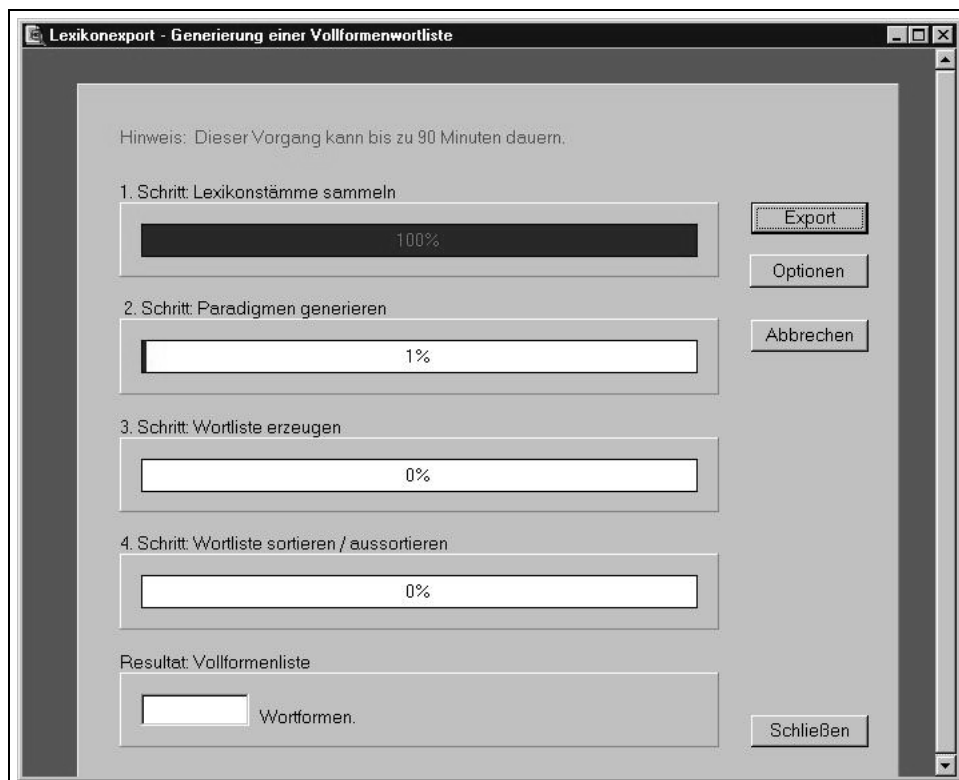
Wortklasse	Anzahl
Substantive	17382
Adjektive	22184
regelmäßige Verben	3792
unregelmäßige Verben	528
Verbzusätze	4253
Sonstige	639
Hochfrequente	397
Eigennamen	1409
Summe	50584

5.3 Export des Morphy-Lexikons

Zunächst sollten Sie die Optionen zum Export aufrufen und dort festlegen, welchen Teil des Lexikons Sie extrahieren möchten:



Hier können Sie wählen, ob Sie nur Stammformen oder alle Vollformen erzeugen möchten. In der Regel werden Sie Vollformen wählen. Daneben können Sie den Export auf einzelne Wortklassen einschränken. Wählen Sie dazu die Option *Einschränkung auf einzelne Wortklassen* und wählen Sie anschließend in der Liste alle benötigten Wortklassen aus. Mit Hilfe dieser Optionen läßt sich z.B. auf einfache Weise ein Lexikon aus Verb- und Adjektivformen zusammenstellen.



Nachdem die Optionen gewählt sind (Standardeinstellung ist Vollformen aller Klassen), starten Sie den Vorgang durch Anklicken von *Export*. Der Export kann je nach verwendetem PC-Typ zwischen 30 und 90 Minuten in Anspruch nehmen. Er ist jedoch jederzeit abbrechbar.

Noch ein Hinweis: Für die Erzeugung eines Vollformen**lexikons** stellt die Exportfunktion nur eine Zwischenstation dar. Denn hier exportieren Sie lediglich eine Vollformen**liste**. Damit aus der Liste ein Lexikon wird, schicken Sie die Liste durch die Morphologische Analyse. Hier können Sie dann auch die Analysestufe nach Ihren Wünschen einstellen.