

Programmierprojekt

Morphosyntaktische Analyse von  
italienisch-deutsch  
gemischtsprachigen Sätzen für SVOX

Sandra Roth  
Im Sandacher 14  
8154 Oberglatt  
Tel.: 01/850 66 40  
E-Mail: saen-d@gmx.ch

7. Juli 2005

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>polySVOX</b>	<b>2</b>
<b>3</b>	<b>Italienische Fonologie</b>	<b>3</b>
3.1	Vokalismus . . . . .	3
3.1.1	Vokallänge . . . . .	3
3.2	Raddoppiamento fonosintattico (fonosyntaktische Verdoppelung) . . .	3
3.2.1	Implementation . . . . .	5
3.3	Akzent . . . . .	6
<b>4</b>	<b>Das italienische Lexikon</b>	<b>7</b>
4.1	Nomenklassen . . . . .	7
4.2	Verben . . . . .	8
<b>5</b>	<b>Intonation / Prosodie des Italienischen</b>	<b>10</b>
<b>6</b>	<b>Phänomene in gemischtsprachigen Sätzen</b>	<b>11</b>
6.1	Deutsche Sätze mit italienischen Einschlässen . . . . .	11
6.1.1	Einschlüsse von ganzen Phrasen . . . . .	11
6.1.2	Einschlüsse mit deutscher Morphologie/Morphosyntax . . . . .	11
6.1.3	Einschlüsse mit italienischer Morphosyntax . . . . .	12
6.1.4	Problem- und Zweifelsfälle . . . . .	12
6.2	Italienische Sätze mit fremdsprachigen Einschlässen . . . . .	13
<b>7</b>	<b>Diskussion und Ausblick</b>	<b>14</b>
	<b>Bibliographie</b>	<b>15</b>
	<b>Anhang A: Testbeispiele</b>	<b>16</b>
	<b>Anhang B: Aufgabenstellung</b>	<b>19</b>

# 1 Einleitung

Im Rahmen dieser Arbeit habe ich hauptsächlich grammatische und phonotaktische Regeln für den italienischen Teil des Sprachsynthesystem polySVOX erstellt. Die Tests wurden einerseits mit dem rein italienischen System gemacht, andererseits mit dem deutsch-italienisch gemischten System, wobei hauptsächlich deutsche Sätze mit italienischen Einschlüssen getestet wurden, aus dem praktischen Grund, dass das italienische Lexikon noch nicht sehr umfangreich ist.

Im nächsten Kapitel versuche ich, grob die Struktur von polySVOX zu erläutern. In den drei darauf folgenden Kapiteln werden zunächst rein italienische Phänomene besprochen, welche auch für Einschlüsse in deutschen Sätzen relevant sein können und natürlich vor allem für das italienische Lexikon und die italienische Grammatik implementiert werden. Im Kapitel 6 werden verschiedene mögliche Beispiele von Einschlüssen vorgestellt. Im letzten Kapitel schliesslich werden offene Fragen und weitere mögliche Implementierungen diskutiert.

## 2 polySVOX

PolySVOX besteht aus mehreren Komponenten auf verschiedenen Ebenen. Auf der Wortebene werden zwei Lexika verwendet, ein Vollformenlexikon (*flex*) und ein Morphemlexikon (*mlex*), in welchen sowohl die Wortkategorie als auch die graphemische und die phonemische Form eines Lexems bzw. eines Morphems eingetragen werden. Die Komponenten (Stamm, Prefix, Suffix etc.) des Morphemlexikons werden über die Wortgrammatik (*wgram*) "zusammengesetzt". Für die Wortanalyse werden sowohl graphische als auch phonetische Two-Level-Rules (Informationen zu den Two-Level-Rules siehe Traber 1995: S.78ff) verwendet. Die graphemischen Two-Level-Rules werden in der Datei *mograph* geschrieben. Zum Beispiel wird dort festgelegt, dass bei gewissen Verben mit dem Konjugationstyp *-care* oder *-gare* ein 'h' vor Verbendungen mit den Vokalen 'e' oder 'i' eingeschoben werden muss (*pregare* - *prego* - *preghi*). Die morphophonetischen Regeln sind in der Datei *mophon* zu finden. Da wird z.B. geregelt, dass die velaren Plosive [k] und [g] vor den Vokalen e und i als Affrikate ([tʃ] und [dʒ]) realisiert werden. Auf Satzebene wird der Satz dann mit Hilfe der Satzgrammatik (*sgram*) geparkt. Als Resultat des Parsing wird ein Syntaxbaum ausgegeben. Aus dem Syntaxbaum werden die Akzentuierung und die Phrasierung abgeleitet. So entsteht die phonologische Repräsentation des ganzen Satzes.

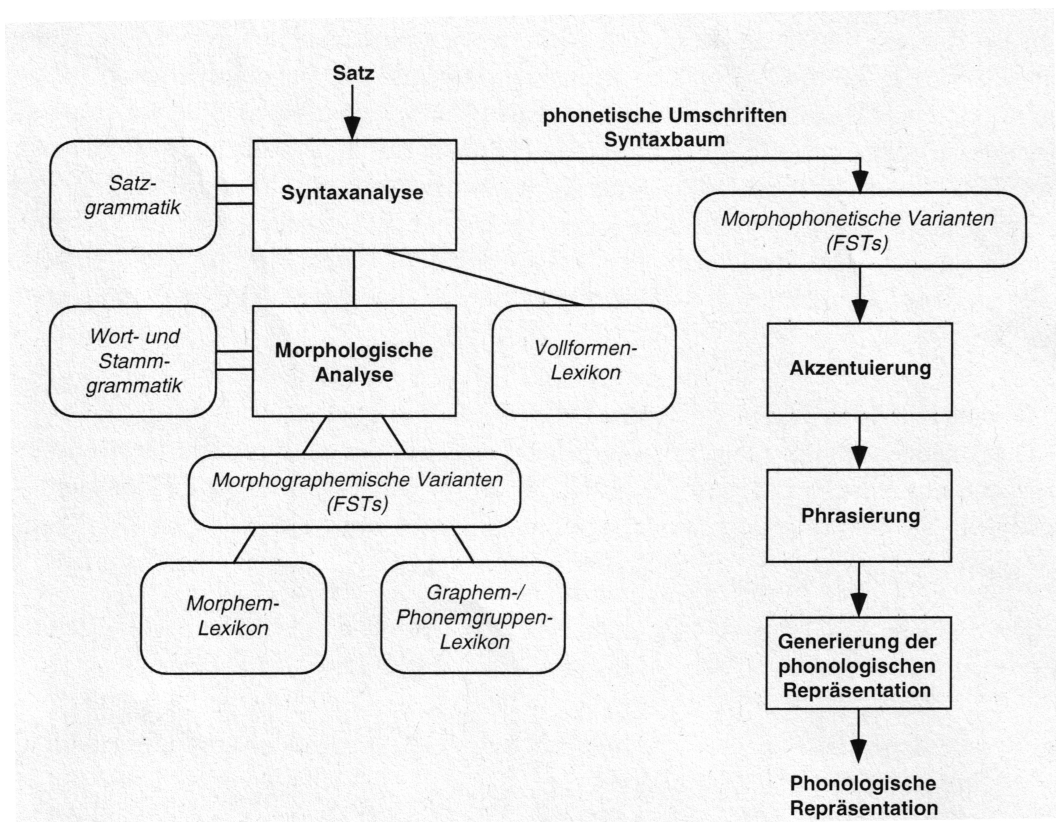
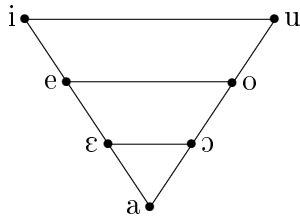


Abbildung 1: Struktur von polySVOX (Pfister et al. 2004)

# 3 Italienische Fonologie

## 3.1 Vokalismus

Die Vokale des Standarditalienischen:



Die Opposition zwischen den offenen Vokalen [ɛ] und [ɔ] und den geschlossenen Vokalen [e] und [o] besteht nur unter Betonung (Minimalpaar: ['venti] 'Zwanzig' vs. ['venti] pl. von *vento* 'Wind'). Ist der Vokal unbetont, wird er geschlossen (Bsp.: [te'le:fono] vs. [telefo'nimo]).

### 3.1.1 Vokallänge

Im Standarditalienischen ist die Länge keine fonologisch relevante Eigenschaft. Ausser den betonten Vokalen in offener Silbe sind alle Vokale des Italienischen kurz. Aufgrund dessen lässt sich folgende Regel aufstellen (vgl. Schmid 1999: S.165):

$$V \rightarrow [+lang] / \overline{[+akzent]} \$$$

Diese Regel wird für die Analyse unbekannter Wörter verwendet, muss aber auch bei Wortbildungen mit Suffixen, bei denen sich der Akzent verschiebt, beachtet werden.

## 3.2 Raddoppiamento fonosintattico (fonosyntaktische Verdoppelung)

Ein wortübergreifendes fonologisches Phänomen im Italienischen ist die fonosyntaktische Verdoppelung (RF = Raddoppiamento fonosintattico).

Die für das heutige Italienisch typische Verdopplung besteht aus einem phonologisch regelmässigen Teil (Verdopplung der Anfangskonsonanten eines Wortes nach betontem Endvokal) und aus einem unregelmässigen Teil (Verdopplung nach einigen unbetont auslautenden Ein- und Mehrsilbern) (Loporcaro 1997).

Die das Raddoppiamento auslösenden Wörter können in die folgenden vier Kategorien aufgeteilt werden:

- a. **alle** mehrsilbigen endbetonten Wörter *andrò* [v:]*ia*
- b. **alle** "starken"<sup>1</sup> einsilbigen Wörter (betont) *sto* [b:]*ene*

<sup>1</sup>Wörter, welche einen Akzent tragen können.

- c. **einige** schwache einsilbige Wörter (unbetont) a [l:]ui  
 (a, da, e, fra, ma, né, o, se, su, tra)
- d. **einige** mehrsilbige, auf der vorletzten Silbe betonten Wörter come [t:]e  
 (come, dove, qualche)

Einige Beispiele für starke einsilbige Wörter der Kategorie b. (Schmid 1999:S.171):

unregelmässige Verben im Präsens	<i>sto, sta, so, sa, fa, do dà, ha, può</i>
Adverben	<i>qui, qua, lì, là, già, più</i>
Pronomen	<i>me, te, sé, ciò, ché, chi</i>
Substantive, Adjektive, Zahlwörter	<i>re, sci, scìa, blu, tre</i>

Tabelle 1: Starke einsilbige Wörter

Der Unterschied zwischen starken und schwachen einsilbigen Wörtern ist allerdings nicht ganz so klar und scheint eher eine terminologische Frage zu sein (vgl. Loporcaro 1997:2f).

Da die unregelmässigen Kategorien endlich sind und nur die aufgelisteten Wörter eine Verdoppelung auslösen, lässt sich das ganze Phänomen relativ gut implementieren.

Folgende Regel beschreibt die regelmässige Verdoppelung (vgl. Schmid 1999: S.170f):

$$C \rightarrow C: / \left[ \begin{array}{l} +vokalisch \\ +akzent \end{array} \right] \# \_$$

Im Gegensatz zu den Kategorien c. und d. sind die ersten beiden Kategorien nach Loporcaro (1997) offen und produktiv und schliessen auch Neologismen und Lehnwörter ein (vgl. Loporcaro 1997: S.1).

In zwei Fällen wird die Verdoppelung blockiert (vgl. Schmid 1999: S. 171):

1. wenn der erste Konsonant des von der Verdoppelung betroffenen Wortes an sich schon lang ist, d.h. die Konsonanten /f/, /ŋ/, /k/, /ts/, /dz/
2. wenn das Wort mit einer Verbindung von /s/ oder /z/ + C beginnt

Beide Fälle können mit der Bedingung [# = \$] für die obige Regel ausgeschlossen werden, denn so muss die Wortgrenze mit der Silbengrenze übereinstimmen.

Ein Beispiel für ein RF innerhalb eines italienischen Einschlusses in einem deutschen Satz ist Nr. (15) (siehe Anhang A, S.16): “Geniessen Sie einen **Caffè Latte** [...]”. *Caffè* ist ein Wort der Kategorie a, also ein mehrsilbiges, endbetontes Wort, welches eine regelmässige Verdoppelung auslöst. *Caffè Latte* wird also [ka'fɛ 'latɛ] ausgesprochen.

Die Verdoppelung kommt auch innerhalb von zusammengesetzten Wörtern vor, beispielsweise wenn ein klitisches Pronomen an einen einsilbigen Imperativ angehängt wird (*dimmi, vattene*) oder nach den Prefixen *sopra-* und *contra-* (*soprattutto, contraddire*). Hier wird die Verdoppelung auch auf der graphemischen Ebene sichtbar.

### 3.2.1 Implementation

Für die regelmässige Verdoppelung kann in den Two-Level-Rules der (Satz-)Akzent auf der letzten Silbe eines Wortes als Auslöser benutzt werden, während für die unregelmässige Verdoppelung ein Trigger eingesetzt werden muss. Bsp. für regelmässiges RF in der Datei `synphon_I`:

```
'b'/'b_b' <=> ('1' | '2' | '3' | '4') '\ ' %Vow_I ']' ' '[' _ ;
's'/'s_s' <=> ('1' | '2' | '3' | '4') '\ ' %Vow_I ']' ' '[' _
              '\ ' 'I' ';' ('0'|'1'|'2'|'3'|'4') '\ ' %Vow_I;
```

Die erste Regel heisst etwa: [b] wird am Anfang eines Wortes ('[') zu [b:], wenn sich im linken Kontext (\_ = Platzhalter) ein betonter ('1-4') Vokal (%Vow\_I) am Ende des vorangehenden Wortes (']') befindet. Der Regel für die Verdoppelung von [s] wird noch hinzugefügt, dass die Verdoppelung nur geschehen darf, wenn ein Vokal folgt (damit keine unmöglichen Konsonantencluster entstehen).

Bsp. für das unregelmässige RF:

```
'k'/'k_k' <=> 'ü' ']' ' '[' _ ;
```

In dieser Regel wird als Kontext nur der Trigger am Wortende verlangt. Der Trigger wird den Wörtern, welche das Raddoppiamento auslösen, in der phonemischen Form angefügt, z.B.

```
INDEF_I (id_qualche,sg,m,adj) "qualche" "k'^u_alkeÜ".
```

Da jene Wörter alle im Vollformenlexikon zu finden und nur auf wenige Wortartklassen verteilt sind, kann man auch nur diese Klassen den Two-Level-Rules übergeben.

Die Verdoppelung (graphemisch und fonemisch) innerhalb der Wörter passiert auf der Wortebene, deshalb werden die Regeln dafür in den Dateien `mophon_I` und `mograph_I` (das ist nötig, damit ein Wort wie *dammelo* überhaupt als Verb mit angehängten Klitika erkannt und dementsprechend auch auf der ersten Silbe betont wird). Da die die Verdoppelung auslösenden Wörter relativ wenige sind, wird wie beim unregelmässigen Raddoppiamento Fonosintattico mit Trigger gearbeitet. In den Two-Level-Rules wird jeweils nach den Triggern für die Verdoppelung gesucht.

### 3.3 Akzent

Das prosodische System des Italienischen besitzt die Eigenschaft des lexikalischen Akzents, d.h. für jedes mehrsilbige Wort muss die Position des Akzents gelernt werden. Der Akzent hat daher auch Unterscheidungscharakter; es lassen sich z.B. folgende Minimalpaare bilden (vgl. Schmid 1999: S.173f): *papa* ~ *papà*, *àncora* ~ *ancora*, *prédico* ~ *predico*.

Das unmarkierte Akzentpattern des Italienischen ist die Betonung der zweitletzten Silbe. 62 % der Wörter des LIP-Korpus (Lessico di frequenza dell'italiano parlato) sind auf der zweitletzten Silbe betont. Daher wird für die unbekanntenen Wörter die Betonung auf der zweitletzten Silbe angenommen.

Es gibt aber trotzdem gewisse Regeln: So kann der Akzent z.B. nicht auf die drittletzte Silbe fallen, wenn die zweitletzte Silbe geschlossen ist. Ausnahmen dieser Regel machen hauptsächlich Ortsnamen (*Tàranto*, *Lèvato*) und einige wenige Nomen (*màndorla*, *pòlizza*).

Bei SVOX ist es so, dass immer der Akzent am weitesten rechts als Hauptakzent angenommen wird. Wenn der Stamm eines Verbs also einen Akzent besitzt, die Endung jedoch auch, wird die Endung hauptbetont.

Bei Wörtern mit drei oder mehr Silben sind auch Nebenakzente vorhanden. Ein Nebenakzent grenzt aber nie direkt an einen Hauptakzent. So ist der Nebenakzent im Adverb [general'mente] beispielsweise auf der ersten Silbe und nicht auf der ursprünglich betonten Silbe des Adjektivs [gene'ra:le] (vgl. Schmid 1999: S.174).



## 4 Das italienische Lexikon

Als Referenz für die phonemische Repräsentation der italienischen Wörter wurde das Pons Kompaktwörterbuch Italienisch-Deutsch verwendet. Dazu ist anzumerken, dass in solchen Wörterbüchern z.B. nur die Grundform von Verben phonetisch notiert sind. Das heisst, dass nicht immer klar ist, ob bei Flektierungen geschlossene Vokale des Stammes zu offenen werden oder ein *i* zu einem Semikonsonanten etc.

### 4.1 Nomenklassen

In älteren italienischen Grammatiken werden meistens drei Flexionsklassen für das Standarditalienisch angenommen, nämlich als erste Klasse die Nomina mit der Endung *-a* (pl. *-e*), als zweite Klasse Nomina mit der Endung *-o* (pl. *-i*) und als dritte Klasse jene mit Endung *-e* (pl. *-i*). Als Ausnahmen werden Nomina wie *virtù* (Pl. *virtù*) oder *textitbraccio* (pl. *braccia*) genannt. D’Achille und Thornton (2003) beschreiben die Flexionsklassen der italienischen Nomina etwas komplexer und gehen dabei nicht davon aus, dass das Genus die grammatischen Kategorien der Nomina bestimmt, sondern die Relation zwischen Singular und Plural. So ergeben sich nach D’Achille und Thornton (2003) folgende Flexionsklassen:

Kl.	Form (sg./pl.)	Beispiel	Genus (mehrheitlich)	Anmerkung
1	-o/-i	libro/libri	m	mano/mani f
2	-a/-e	casa/case	f	
3	-e/-i	fiore/fiori, siepe/siepi, cantante/cantanti	m/f	
4	-a/-i	poeta/poeti	m	ala/ali f, arma/armi f
5	-o/-a	uovo/uova	sg. m, pl. f	
6	unveränderl. Pl.	re, gru, città, virtù, crisi, film, caffè	m/f	
7	-o/-ora	prato/pratora	sg. m, pl. f	nur im Altitalienischen

Tabelle 2: Flexionsklassen italienischer Substantive

Die Klassen 4 und 5 sind im heutigen Italienisch abgeschlossene Klassen, das heisst, es kommen keine neuen Wörter dieser Klassen mehr zum Lexikon hinzu. Produktiv ist nur noch das Suffix *-ista*, mit welchem neue Nomina der 4. Klasse gebildet werden können. Wie in der Tabelle vermerkt ist die Flexion der Klasse 7 im heutigen Italienisch nicht mehr vorhanden. Im Lexikon von SVOX sind diese Flexionsklassen bis jetzt noch nicht ganz so umgesetzt; momentan gibt es folgende Klassen:

Klasse	Bsp.	Anmerkungen
o	arancia/-e, ragazzo/-i, poeta/-i	
of	mano, -i	
co	amico/-i/-a/- he, parco/-hi	
io	commedia/-e, figlio/figli	diese Klasse braucht es, da Diphthonge nur zusammen (also in der Endung) geschrieben werden können
uo	lingua/-e	siehe Anm. -io
e	paese/-i, siepe/-i	
vo	braccio/-a	
null	caffè, città, ragù	

Tabelle 3: Flexionsklassen in SVOX

Unregelmässige Deklinationen wie *uomo-uomini* werden direkt als Vollformen (NF\_I) im Morphologielexikon eingetragen.

## 4.2 Verben

Die Verben werden in SVOX in 7 “regelmässige” und einige unregelmässige Klassen unterteilt. Regelmässig sind folgende Paradigmen:

Klasse	Beispiel	Anmerkungen
-are	pescare	
-ere	appartenere, piacere	
-ire	partire	
-iare	macchiare	diese Klasse wird gebraucht, weil sich der Diphthong nicht trennen lässt
-cere	credere	wird nur für nicht betonte Infinitive von <i>-ere</i> -Verben
-gnere	spegnere	<i>-ere</i> -Klasse kann nicht verwendet werden, da sich der palatale Nasal [ʎ] z.B. in der ersten Person in einen velaren Nasal gefolgt von einem g umwandelt ['spɛŋgo].
-isc	finire, preferire	

Tabelle 4: It. Verbklassen in SVOX

Unregelmässige Verben wie *andare, fare, dire* etc. sind teilweise in eigenen Paradigmen aufgelistet, für gewisse Tempi oder Modi können sie aber regelmässig dekliniert werden (Bsp.: Imperfekt von *andare*: *andavo, andavi, andava, andavamo, ...*), oder stimmen mit Paradigmen von anderen unregelmässigen Verben überein. So muss nicht das vollständige Paradigma eines unregelmässigen Verbs ins Morphologielexikon eingetragen werden, sondern es sind nur die spezifisch unregelmässigen Endungen einzutragen. Die Verstämme werden aufgrund der unregelmässigen Verben mit

den Features Klasse, Modus, Tempus und Person ins Lexikon eingetragen.

Beispiel eines Eintrags eines regelmässigen Verbstammes:

```
VS_I (id_appartenere,ere,?,?,?,?) "appart+" "ap_part+"
```

Beispiel eines Eintrags eines unregelmässigen Verbes, bei dem sowohl die Deklinationsklasse, als auch der Stamm wechselt:

```
VS_I (id_andare,andare,ind,pres,sg,?) "v+" "v+"  
VS_I (id_andare,are,ind,pres,pl,p1) "and+" "and+"
```

## 5 Intonation / Prosodie des Italienischen

Der Satzakzent fällt als “Default” auf den Hauptakzent der letzten (intonativen) Phrase des Satzes. (vgl. Schmid 1999: S.113). Im Italienischen gibt es drei Grundintonationstypen (vgl. Schmid 1999: S. 175f):

<b>Intonation</b>	<b>Funktion</b>
sinkend	terminal (abschliessend)
steigend	fragend
gleich bleibend	Kontinuität anzeigend

Ansteigende Intonation (auf der letzten Silbe) das einzige Mittel um bei “Ja-Nein-Fragesätzen” die interrogative Modalität anzuzeigen. Bsp.: *Gianni è venuto?*.

Abgesehen von den drei Grundtypen gibt es noch viele weitere Intonationstypen, unter anderem:

<b>Intonation</b>	<b>Funktion</b>
sinkend-steigend	Zweifel, Überraschung
steigend-sinkend	“energische” Behauptungen

## 6 Phänomene in gemischtsprachigen Sätzen

### 6.1 Deutsche Sätze mit italienischen Einschlüssen

Die Beispielsätze, welche auch zum Testen verwendet wurden, sind im Anhang A ab S.16 zu finden. Der grösste Teil der fremdsprachigen Einschlüsse, seien es nun italienische Einschlüsse im Deutschen oder deutsche, englische oder französische Einschlüsse in italienischen Sätzen, besteht aus einzelnen Nomen oder Eigennamen. Meistens wird das grammatische Geschlecht des Nomens seinem Äquivalent in der jeweiligen “Basissprache” angepasst:

- (10) Rein optisch spricht alles für das **Bel Paese**. (it. ‘il paese’ vs. dt. ‘das Land’)
- (33) Che cosa univa la mia infanzia alla **Deutsch-Ostafrika**? (it. ‘Africa’ (f.) vs. dt. ‘Afrika’ (n.))

Allerdings kann man auch gegenteilige Beispiele finden. So ist im Satz (4) “*die Piazza*” feminin, auch wenn das deutsche Äquivalent dazu maskulin ist (‘der Platz’).

#### 6.1.1 Einschlüsse von ganzen Phrasen

(10) bis (14) sind Beispiele für nominale Einschlüsse mit attributivem Adjektiv, wobei die ganze Phrase durch deutsche Artikel oder Präpositionen meistens zu deutschen Nominalphrasen werden. Im Satz (14) wird sogar eine komplette italienische Nominalphrase (“*la dolce vita*”) eingeschlossen.

Der Satz (15) ist ein Beispiel für einen Einschluss eines zusammengesetzten italienischen Substantivs. In solchen Fällen sollte die italienische Phrasenbetonung angewendet werden. Das Kompositum wird nicht wie in deutschen Phrasen links sondern rechts betont (der Hauptakzent liegt also auf *latte* und nicht auf *caffè*).

#### 6.1.2 Einschlüsse mit deutscher Morphologie/Morphosyntax

Wie schon in Kapitel 6.1 angedeutet, wechseln deutsche Artikel von italienische Einschlüssen oft das grammatikalische Geschlecht. Ein schönes Beispiel ist Nummer (19), bei welchem man die Femininmarkierung in den italienischen Wörtern deutlich an den Endungen sieht (*insalata mista*), dem zum Trotz aber dennoch der maskuline Artikel (des maskulinen deutschen Äquivalents) vorangestellt wird. Des Weiteren können auch gemischtsprachige Komposita gebildet werden wie die Beispiele (20) (‘Vivace-Satz’), (22) (‘Duce-Enkelin’) und (1) (‘Italianità-Rating’) beweisen.

Ein Beispiel für einen Einschluss mit deutscher Morphologie bietet der Satz (23). An den italienischen Eigennamen *Machiavelli* wird das deutsche Suffix *-ismus* angehängt. Ein ähnlicher Fall ist das “Bologneser ragu” von Satz (25). Wahrscheinlich

wird das italienische Adjektiv ('bolognese') direkt mit der deutschen Endung -r versehen.

Deutsche Morphologie findet sich auch im Satz (24). In diesem Fall wird ebenfalls italienischen Eigennamen ein deutsches Genitiv-s angehängt (*Dantes Divina Commedia und Boccaccios Decameron*).

Wie man am Satz (21) sieht, kann man bei gebräuchlicheren Einschlüssen anstelle des italienischen Plurals aus dem gesamten italienischen Wort (Stamm und Endung) auch einen s-Plural des deutschen bilden.

### 6.1.3 Einschlüsse mit italienischer Morphosyntax

Im Beispiel (25) findet sich auch ein Fall, in welchem die italienische Morphologie auf die deutsche einwirkt: *lasagne* ist im Italienischen eine Pluralform, die jedoch im Deutschen meist als Singular verwendet wird; in diesem Satz jedoch wird das Wort explizit als Plural verstanden und das deutsche Verb entsprechend angepasst (" [...] *sind lasagne al forno*").

Ebenfalls ein eher seltenes Phänomen trifft man in Satz (6) an: das italienische Adjektiv wird, wie in der italienischen Syntax, dem deutschen Substantiv nachgestellt.

### 6.1.4 Problem- und Zweifelsfälle

Im Satz (13) kommt das Wort *Soloinstrumente* vor. Die Frage ist nun, ob *Solo* als deutsches oder als italienisches Wort behandelt, also im deutschen Lexikon aufgeführt werden soll. Zum einen wird das *s* im Deutschen stimmhaft ausgesprochen, während es im Italienischen stimmlos ist. Interessant ist auch die Substantivierung des Adjektivs im Deutschen, welche es im Italienischen nicht gibt (dt. 'Solo', it. 'assolo'). Dieselbe Frage kann man sich bei Lexemen wie *Cappucino* oder *Spaghetti* stellen.

Ein Problem für die Inklusionsgrammatik stellt das Beispiel (6) mit dem nachgestellten Adjektiv dar. Wahrscheinlich müsste es so interpretiert werden, dass *Lieder tedeschi* als eine italienische Phrase verstanden wird, das deutsche Nomen *Lieder* also auf ein italienisches abgebildet und anschliessend die ganze (Teil-)Phrase mit Adjektiv von Italienisch nach Deutsch zurückgemappt wird. Da die deutsche Grammatik nachgestellte Adjektive nicht zulässt und in diesem Fall einfach das italienische Adjektiv auf ein deutsches abgebildet wird, kann momentan kein korrekter Syntaxbaum erstellt werden.

Bei einem System wie SVOX sind Ambiguitäten natürlich ein Problem. So ist es im

Satz (8) beispielsweise nicht klar, wie das Numerale ausgesprochen werden soll; *zero* könnte gerade so gut als englisches wie italienisches Numerale interpretiert werden. In diesem Beispiel kann ein menschlicher Leser die Sprache des Einschlusses durch den Kontext entscheiden (der Satz stammt aus einem Zeitungsartikel über Tessinerstädte). In anderen Zusammenhängen kann es aber selbst für menschliche Leser schwierig sein, zu desambiguieren.

## 6.2 Italienische Sätze mit fremdsprachigen Einschlüssen

Obwohl ich mich hauptsächlich darauf konzentriert habe, deutsche Einschlüsse in italienischen Sätzen zu finden, habe ich schlussendlich mehr Beispiele mit englischen oder französischen Einschlüssen zusammengetragen. Somit lassen sich auch nicht viele verschiedene Einschlüsse finden. Auch in diesen Beispielen bestehen die meisten Einschlüsse aus einzelnen Substantiven. Einziges Beispiel eines deutschen Einschlusses mit italienischer Morphologie unter den Testsätzen ist Satz (40). Dem deutschen Eigennamen *Jung* wird der italienische Adjektivsuffix *-iano* angehängt (und das *h* zur korrekten Aussprache des *g* eingefügt).

Der Satz (60) mit französischem Verb-Einschluss ist in morphosyntaktischer Hinsicht ebenfalls sehr interessant. Es werden nämlich insofern italienische Grammatikregeln verwendet, als dass im Französischen obligatorisch ein Subjektspronomen verlangt wird, während diese Pronomen im Italienischen auch weggelassen werden können. In dem kurzen Satz “Ma glissons” (dem Buch *La misteriosa fiamma della regina Loana* von Umberto Eco entnommen) wird somit das französische finite Verb auf ein italienisches gemappt.

## 7 Diskussion und Ausblick

Wie in Kapitel 4 bereits erwähnt, ist es nicht einfach, für die Lexikoneinträge ein Wörterbuch zu finden, welches die phonetische Form eines Wortes eindeutig wiedergibt, also z.B. geschlossene und offene Vokale und Diphthonge transkribiert.

Nicht ganz einfach ist es, eine gute Lösung für die Konjugation der Verben zu finden (siehe Kapitel 4.2). Natürlich könnte man die unregelmässigen Verben ins Vollformenlexikon eintragen. Dann besteht jedoch keine Möglichkeit mehr, morphologische Operationen durchzuführen wie Suffixe anhängen oder die Verben zu substantivieren u.ä.

Eine noch offene Frage ist, wie die Aussprache von englischen und deutschen Einschlässen im Italienischen mit Lauten wie [h], [ɪ] oder [ə] realisiert werden soll. Werden solche Laute von (italienischen) Sprechern ausgesprochen, bzw. verstanden, wenn sie “korrekt” ausgesprochen werden? Die Annahme ist, dass sie von den meisten Sprechern nicht realisiert werden. Schmid (1999: S.140) schreibt, dass der Konsonant [h] vor allem bei gebräuchlichen Wörtern (wie *hobby* oder *hamburger*), welche schon ins italienische Lexikon eingegangen sind, nicht ausgesprochen wird. Im Wörterbuch “Zingarelli” findet man für das Wort *hobby* die fonetische Umschrift [ˈɔbːi]; im Englischen wird das Wort als [ˈhɒbi] realisiert. Nicht nur das h sondern alle im Italienischen nicht gebräuchlichen Phoneme werden ans italienische Inventar angepasst und die geschriebenen zwei <b> werden nach italienischen “Graphem-zu-Phonem-Konventionen” auch als langes b ausgesprochen. Momentan ist es so, dass das System die Einschlässe mit dem Phoneminventar der jeweiligen Sprache realisiert, was bei Einschlässen von ganzen Sätzen oder längeren Phrasen sicher die beste Lösung ist, für kürzere Einschlässe aber noch untersucht werden müsste.

Momentan gibt es nur eine terminale Satzintonation. Man könnte z.B. die Intonation für Fragesätze des Italienischen (siehe Kapitel 5) implementieren.



## Bibliographie

- D'Achille, Paolo und Anna M. Thornton (2003): *La flessione del nome dall'italiano antico all'italiano contemporaneo*, in Nicoletta Maraschio (Hg.) *Italia linguistica anno mille, Italia linguistica anno duemila : Atti del XXXIV Congresso SLI (Firenze 2000)*, Rom: Bulzoni, S. 211–230.
- Loporcaro, Michele (1997): *L'origine del raddoppiamento fonosintattico. Saggio di fonologia diacronica romanza*, Basel und Tübingen: A. Francke Verlag.
- Pfister, Beat, Hans-Peter Hutter und Christof Traber (2004): *Sprachverarbeitung II. Vorlesungsskript für das Sommersemester 2004*, Departement ITET, ETH Zürich.
- Schmid, Stephan (1999): *Fonetica e fonologia dell'italiano*, Turin: Paravia Scriptorium.
- Traber, Christof (1995): *SVOX: The Implementations of a Text-to-Speech System for German*, Dissertation, ETH Zürich.

# Anhang A

## Testbeispiele

Die meisten der folgenden Sätze sind Zeitungen entnommen, manche aus Texten aus dem Internet und einige auch aus der Literatur.

### Deutsche Sätze mit italienischen Einschlüssen

- (1) Tessiner Städte im **Italianità** Rating.
- (2) Erst in der Zeit des “**Risorgimento**” konnte eine nationale Einigung erzielt werden, vor allem durch die Freiwilligenverbände unter **Giuseppe Garibaldi**.
- (3) War das “**Risorgimento**”, also der als “Wiedererstehung” bezeichnete Prozess der liberalen Nationsbildung Italiens, ein mittel- und südeuropäisches Phänomen?
- (4) Grüppchen schnatternder **Ragazzi** mit **Telefonini** auf der **Piazza**.
- (5) Aber eine **Svizzera**, die sich äusserst mediterran gibt.
- (6) Was er nicht gemeint haben kann, sind die “Lieder **tedeschi**”, die man nun eben dort an der Adria hören wird.
- (7) Das Wort “**macchiato**” ist übrigens mit dem deutschen Wort “Makel” verwandt und bedeutet “befleckt”.
- (8) Aber leider auch eines, das **zero** Feingefühl vermittelt.
- (9) Eine hochästhetische Photographie: An einer langen Tafel sitzen zwölf junge Frauen und ein Jüngling, in Designerkleider gewandet, die Finger manieristisch gespreizt, in artistisches **Chiaroscuro** getaucht.
- (10) Rein optisch spricht alles für das **Bel Paese**.
- (11) In den Gassen wabert der Geruch von **Espresso** und **Pollo arrosto**.
- (12) Wohl kaum ein Heißgetränk wird zur Zeit so häufig nachgefragt wie **Latte Macchiato**.
- (13) Das ‘**Divertimento giocoso**’ von Harald Genzmer sieht als dem Streichorchester gegenübergestellte Soloinstrumente zwei Holzbläser vor.
- (14) In einer solchen Atmosphäre regt sich in uns unweigerlich das Bedürfnis, unsere deutsche Identität abzustreifen und die Illusion von “**la dolce vita**” und “**bella Italia**” nicht durch falsche Aussprache all der Köstlichkeiten auf der Speisekarte frühzeitig zerplatzen zu lassen.
- (15) Geniessen Sie einen **Caffè Latte** oder eine feine italienische Spezialität im Salon Rouge des Landesmuseums.
- (16) “**Pranzo al sacco**” nannte sich die mit Essen und Trinken vollgepackte Tasche.
- (17) Diesem besonderen Charakter hat “**La dolce lingua**” angemessenen Raum gelassen, vom “**culto delle tre corone**”, insbesondere im entscheidenden Augenblick der Kodifizierung im sechzehnten Jahrhundert bis hin zu den Instrumenten der Norm, den Grammatik- und den Wörterbüchern, allen voran

das ruhmreiche **Vocabolario della Crusca**.

- (18) Der **Giro d'Italia** gilt nach der Tour de France als zweitbedeutendstes Radrennen der Welt.
- (19) Sie bestellt *einen* **Insalata mista** und die überbackenen **Spinat-Gnocchi**.
- (20) Das Werk setzt einem tänzerisch bewegten **Scherzo** harmonische Querstände bei und schließt mit einem schwungvollen **Vivace-Satz**.
- (21) Es ist im Deutschen aber ebenso erlaubt, “**Cappuccinos**” und “**Espressos**” zu sagen.
- (22) Tatsächlich vertritt die **Duce-Enkelin** [...].
- (23) Solchem **Machiavellismus** hat sich nun aber Fini [...] entschieden widersetzt. h
- (24) **Dantes Divina Commedia** und **Boccaccios Decameron** sind nur zwei der Werke, welche in der Ausstellung zu bewundern sind.
- (25) Typischer Ausdruck der saftig-kräftigen Regionalküche sind **lasagne al forno**: grosse Nudelblätter, schichtweise und abwechselnd mit **Bologneser ragù** und Bechamelsauce übergossen und mit Parmesan überbacken.

#### Italienische Sätze mit fremdsprachigen Einschlüssen

- (26) [...] scherza Renzo Rosso, **patron** della **Diesel**.
- (27) Questa glassa opaca al cioccolato è tipica della torta **Sacher**.
- (28) Negli ultimi 150 anni abbiamo avuto numerose visioni del mondo, prova questa che la **Weltanschauung** è merce screditata [...].
- (29) Si concluse in Europa l'8 maggio 1945 con la resa incondizionata del Terzo **Reich**.
- (30) La loro tattica della **Blitzkrieg** riuscì a sconfiggere i francesi e le armate britanniche in Francia.
- (31) Ciò avviene soprattutto in testi fittizi che stabiliscono una propria **Textwelt**.
- (32) Nella linguistica testuale si distinguono a livello globale il **Weltwissen** e le conoscenze linguistiche.
- (33) Che cosa univa la mia infanzia alla **Deutsch-Ostafrika**?
- (34) Come dice quella barzelletta di Gianni sull'**Alzheimer**?
- (35) Invece i modi per dire “**ich liebe dich**” sono tanti.
- (36) In un testo “**Er sagte...**” è un'espressione che non dice molto del personaggio, delle sue caratteristiche, della situazione.
- (37) Qui ai muri c'erano le stampe tedesche, di fattura molto precisa, **Zur Geschichte der Kostüme**.
- (38) **Vorsicht!** Ora casca, poverino!
- (39) **Was?** Ferito come?
- (40) Come esperto **junghiano** di fama internazionale insegnava a Roma.

- (41) Ma anche il **rock** e i suoi subgeneri preferiscono raggiungere l'ascoltatore con un messaggio intelligibile e diretto.
- (42) Lo **slogan** pubblicitario ha la funzione di dirigere.
- (43) Sussiste fra l'artista e l'ascoltatore (o **fan**).
- (44) La nazionale svizzera di **hockey** su ghiaccio si è qualificata per il torneo olimpico in programma tra un anno esatto a Torino.
- (45) Un concetto chiaro, che mostra quale sarà una delle tendenze di questo e dei prossimi anni: trasformare il telefonino nell'erede del **walkman**.
- (46) Nel decennio successivo il locale era frequentato da celebri **gangster** londinesi.
- (47) E qui prendevi il tuo **whisky** serale.
- (48) E alla fine i **weblog** entrarono alla Casa Bianca. Mentre nelle aule giudiziarie e nei siti specializzati si discute se i **blogger** siano o meno giornalisti, la realtà procede più veloce delle parole.
- (49) **Working poor**: un fenomeno e le sue cause
- (50) In testi di canzone dialettali o **slang** per ascoltatori l'informatività sarebbe di terzo grado.
- (51) Ci sono bar e ristoranti che hanno scoperto il **dehòr** invernale [...].
- (52) Per esempio, martedì scorso Beyoncé è arrivata al **defilé** di Marc Jacobs con un'ora e quaranta di ritardo.
- (53) Come potevo pretendere di condurra la mia **recherche** con lo sfintere?
- (54) Sfrutta la carta, visto che le **madeleines** non ti dicono niente.
- (55) In francese si usa il **vous** anche quando si va a letto.
- (56) Durante il periodo di carnevale i volontari di **Nez Rouge** Ticino hanno effettuato complessivamente 97 interventi per accompagnare a casa 216 persone.
- (57) [...]ebbene, sono sì giunte tre vittorie, ma l'"**en-plein**" non ha dissipato tutti i dubbi sulla rinnovata selezione di Ralph Krueger[...].
- (58) Offre tenera la digeribile bevanda a un avo sorridente, **ancien régime**, ancora vestito di **culottes**.
- (59) [...] per vedere altre scatole raffigurate sulle scatole, **en abîme**, a scatole cinesi a matrioska. (ibidem, p. 123)
- (60) Ma **glissons**.
- (61) Ha il demonio in corpo, mio figlio, capite, **sacré nom de Dieu!**

# Anhang B

## Aufgabenstellung

Sommersemester 2005

### SEMESTERARBEIT

für

Sandra Roth

Betreuer: H. Romsdorfer ETZ D97.5

Stellvertreter: T. Kaufmann ETZ D97.7

---

Ausgabe: 7. Februar 2005

Abgabe: 7. Juli 2005

---

## Morpho-syntaktische Analyse von Italienisch-Deutsch gemischtsprachigen Sätzen

---

### Einleitung

Das in der Gruppe für Sprachverarbeitung zur Zeit entstehende polyglotte Sprachsynthesystem polySVOX erlaubt es, gemichsprachige Sätze vom Computer aussprechen zu lassen. In einem ersten Verarbeitungsschritt werden diese Sätze vom Sprachsynthesystem morphologisch und syntaktisch analysiert. Für diese Analysen werden für jede Sprache sowohl ein Vollformen-Lexikon als auch ein Morphem-Lexikon<sup>1</sup> verwendet (siehe [?], Kap. 3). Diese Lexika enthalten neben Einträgen in graphemischer und phonetischer Form auch die zugehörige grammatikalische Information unter Verwendung des DCG-Formalismus<sup>2</sup>.

Die grammatikalische Einbettung fremdsprachiger Wörter in einen Satz erfolgt mittels so genannter Einschlussregeln, die für bestimmte fremdsprachige Wort-Kategorien definieren, als welche Wort-Kategorien in der Basissprache des Satzes sie verwendet werden dürfen.

---

<sup>1</sup>Als Morphem bezeichnet man den kleinsten bedeutungstragenden Teil eines Wortes.

<sup>2</sup>Der Definite Clause Grammar (DCG) - Formalismus dient zur Beschreibung kontextsensitiver Sprachen unter Verwendung eines kontextfreien "Skeletts" mit zusätzlichen einschränkenden Bedingungen (vergl. [?], Kap. 7.4).

## Problemstellung

Die Aufgabe der morpho-syntaktischen Analysestufe ist es, einerseits für jedes Wort des Eingabetextes die korrekte Lautsequenz inklusive Sprachzugehörigkeit und Wortbetonung in kanonischer Form (d.h. die standardisierte Aussprache des in Isolation gesprochenen Wortes) zu ermitteln, und andererseits die syntaktische Struktur des Satzes für die nachfolgende phonologische Transformationsstufe zu generieren.

Die Palette fremdsprachiger Einschlüsse reicht vom Einschluss einzelner Morpheme (wie z.B. in “**upgedatet**”) über Einzelwörter (z.B. in “Der **Duce** war ...”) bis zum Einschluss von längeren Wortgruppen oder auch Gliedsätzen (wie etwa in “Zuerst genießt man das **dolce vita** im **Ristorante Lorenzini**. Danach geht die **Party** im **Du Théâtre** zum **70s-, 80s-** und **House-Sound** von **Stefano** ab.” oder in “... ungewohnte Dimensionen, die schnell Vergleiche mit einer **piazza d’armi** evozierten und eine grüne Stadtparlamentarierin seufzen ließen, **’è prosa, quello vecchio era poesia’**”).

Damit das Sprachsynthesystem solche Einschlüsse korrekt verarbeiten kann, müssen zuerst diese Einschlüsse innerhalb ihrer jeweiligen Sprache, in diesem Fall Italienisch, morphologisch und syntaktisch analysiert werden können. Danach müssen sie über Mischregeln in die Basissprache des Satzes eingebettet werden.

Die Wortbetonung wird primär über spezielle Betonungsmarkierungen der Lexikoneinträge ermittelt. Zusätzlich kann man für komplexere Wortzusammensetzungen Betonungsregeln definieren, welche die lexikalischen Markierungen geeignet verschieben können.

Da es für die Betonung italienischer Einschlüsse in deutschen Sätzen keine definierten Richtlinien gibt, muss man eigene entwerfen. Denkbar ist, dass längere italienische Einschlüsse der italienischen Standardbetonung folgen, während kurze Einschlüsse eher von der deutschen Betonung geprägt sind.

Bei der Entwicklung der Wort- und Satzgrammatik, bzw. den Betonungsregeln haben sich folgende grundsätzliche Richtlinien als sinnvoll erwiesen:

- Modularer Aufbau der Wortgrammatik durch Trennung von Suffixen und Stämmen (z.B. Verb => Verbpräfix + Verbstamm + Verbsuffix).
- Unterscheidung von Homographen (Wörter mit gleicher Schreibweise) mit unterschiedlicher Aussprache oder Betonung muss über die Syntax möglich sein (z.B. “crogioolo” wird als Verb [krodʒolo] und als Nomen [kroʒo:lo] gesprochen).
- Konstituenten müssen über einfache 1-1 Einschlussregeln in die deutsche Grammatik eingebettet werden können. Voraussetzung dafür ist der oben genannte modulare Aufbau der Grammatik.

## Aufgaben

In Rahmen dieser Semesterarbeit sind die folgenden Aufgaben zu lösen:

1. Einarbeitung in den DCG-Formalismus und das Lexikon-Format von SVOX anhand folgender Literatur ([?], [?]).

2. Sammlung von gemischtsprachigen Beispielsätzen mit italienischen Einschläüssen in deutschen Sätzen und umgekehrt. Anhand dieser Beispielsätze sollen die verschiedenen morphologischen, syntaktischen, aber auch betonungsrelevanten Phänomene bei Deutsch-Italienisch gemischtsprachigen Sätzen untersucht werden.
3. Einfügen der für die Beispielsätze notwendigen Einträge ins Lexikon. Dabei kann das bestehende deutsche Lexikon verwendet werden. Für Italienisch ist ein rudimentäres Lexikon mit wenigen, beispielhaften Einträgen vorhanden.
4. Erstellen der für die Beispielsätze nötigen italienischen Wort- und Satzgrammatikregeln. Dazu gibt es ebenfalls eine rudimentäre Grammatik mit wenigen, beispielhaften Regeln.
5. Erarbeitung der notwendigen Betonungsregeln für Italienisch, im speziellen für italienische Einschläüsse in deutschen Sätzen.
6. Testen der erstellten Regeln mit Hilfe der gesammelten Beispielsätze.
7. Die ausgeführten Arbeiten und die erhaltenen Resultate sind in einem Bericht zu dokumentieren, der in zwei Exemplaren abzugeben ist, wovon eines Eigentum des Instituts bleibt.

### Literaturverzeichnis

- [1] C. Traber. *SVOX: The Implementation of a Text-to-Speech System for German*. PhD thesis, No. 11064, Computer Engineering and Networks Laboratory, ETH Zurich (TIK-Schriftenreihe Nr. 7, ISBN 3 7281 2239 4), March 1995.
- [2] B. Pfister, H.-P. Hutter und C. Traber. *Sprachverarbeitung II*. Vorlesungsskript für das Sommersemester 2003, Departement ITET, ETH Zürich, 2003.

Zürich, den 7. Februar 2005

Prof. Dr. L. Thiele