

**Pluridimensionale Varietätenlinguistik in der Datenbank**

**Entstehung eines Atlases zur morphosyntaktischen Variation im Kalabresischen (ASICA)**

**Abstract**

The project ASICA focuses on the southern Italian dialect spoken in Calabria. Its main purpose is to visualize linguistic analysis drawn from a database in form of online-published maps of Calabria. The database comprises more than 400.000 words collected in interviews with native-speakers from Calabria, approximately half of them living in Calabria, the other half living in Germany. The text of the interviews has been transcribed using the program "Praat" and finally imported into a MySQL-Database. Using apt SQL-queries, specific information on morpho-syntactic structures is extracted from the database. Finally the maps are built based on this information and published on a webpage.

**1. Syntax und Variation – ein Desiderat in der italienischen Dialektologie**

Syntax und ihre pluridimensionale Variation in kalabresischen Dialekten sowie deren Kartierung im Rahmen eines Online-Atlases ist das Ziel eines im Jahr 2004 begonnenen Münchener Forschungsprojekts (ASICA – *Atlante sintattico della Calabria*)<sup>1</sup>.

Die traditionelle italienische Dialektologie ist bekanntlich dem Ideal des bodenständigen Sprechers und seinem archaischen Dialekt stark verhaftet (vgl. Jaberg/Jud 1928). Diese Haltung wird besonders in den Bemerkungen des wichtigen Kommentarbands *Der Sprachatlas als Forschungsinstrument* zum AIS offenkundig und lebt über viele Jahrzehnte in den bis heute prägenden Forschungen Rohlfs zu den süditalienischen Dialekten weiter.<sup>2</sup>

Neuere Werke der italienischen Dialektologie, wie der sizilianische Sprachatlas ALS (*Atlante Linguistico della Sicilia*) fokussieren dagegen pluridimensional gerichtete Perspektiven (vgl. Ruffino 1995).

---

<sup>1</sup>Eine ausführliche Projektskizze findet sich auf der AsiCa-Homepage:

<http://asica.gwi.uni-muenchen.de/index.php?impressum=1>

<sup>2</sup>Einen hervorragenden Überblick zur Ausgrenzung pluridimensionaler Sichtweisen in der italienischen Dialektologie verschaffen Krefeld 2005 u. 2006.

Das Münchner Projekt ASiCA dient der Erhebung markanter Phänomene kalabresischer Syntax, die in der romanischen Sprachwissenschaft seit Meyer-Lübke ihren festen Platz innehaben, jedoch bislang nur im Rahmen der diatopischen Variation betrachtet wurden. Ein Ziel des ASiCA ist es, diese vermeintlich wohlbekanntesten syntaktischen Besonderheiten in bezug auf ihre Variation nicht nur diatopisch, sondern auch syntopisch zu untersuchen (vgl. Krefeld 2007b, Salminger 2002). Dabei spielen Faktoren wie Alter, Bildungsgrad, Migration bei der Informantenauswahl eine wichtige Rolle.

Um dem Anspruch der Erhebung von Variation in einer Nicht-Standard-Varietät gerecht zu werden, ist die Wahl der Interviewtechnik von herausragender Bedeutung. Das Dialektkorpus des ASiCA entstammt einem semispontanen Interviewteil, in dem hauptsächlich biographische Informationen wie Kindheit, Heirat, Leben im Dorf und häufig ethnographisch relevante Themen, z.B. Ernte, Ackerbau, Webhandwerk und kulinarische Spezialitäten im Vordergrund stehen, sowie einem Syntaxfragebogen, der die für die süditalienischen Dialekte wichtigsten Syntaxphänomene<sup>3</sup> fokussiert.

Viele der syntaktischen Muster im elizitierten Material finden sich im Spontankorpus wieder, sodass basilektale Strukturen zuverlässig sowohl qualitativ als auch quantitativ analysiert werden können.<sup>4</sup>

Die Ergebnisse dieser mannigfaltigen Analysen werden schließlich in Form eines Online-Atlas kartiert. Dies hat den Vorteil, dass auch zu einem späteren Zeitpunkt zusätzlich hinzugefügte Ortspunkte problemlos in die Visualisierung miteinbezogen werden können. Desweiteren sind die Audiofiles ebenfalls online zugänglich und eng an die Transkription gekoppelt. Darüber hinaus hat der Benutzer die Möglichkeit, sich Analysen seiner Wahl anzeigen zu lassen, z.B. die Stellung der Possessiva oder die Serialisierungstypen nach *verba movendi*. Diese Optionen machen ASiCA zu einem flexiblen, immer verfügbaren und beliebig erweiterbaren Forschungsinstrument.

---

<sup>3</sup>Innerhalb des Explorationsgebiets verlaufen markante Isoglossengrenzen, welche die nord- und mittelkalabresischen Dialekte von südlichen abtrennen. Zum Verlauf der Issglossengrenzen vgl. u.a. Rohlf's 1949 u. 1985, Radtke 1988, Trumper 1997, Vincent 1997

<sup>4</sup>Die unterschiedlichen Interviewtechniken ermöglichen zudem die kritische Betrachtung von sprachlichen Interferenzen, welche häufig im elizitierten Material auftreten. Die Strukturen des italienischen Stimulus werden vor allem von wenig gebildeten Sprechern bei der Übersetzung nicht adäquat in den jeweiligen Dialekt übersetzt, sondern lediglich lautlich angepasst. Die große Menge an spontansprachlichen Daten relativiert jedoch meist derartige Lücken. Andererseits können in der freien Rede eher selten zu findende morphosyntaktische Aspekte, wie z.B. die Modusabfolge im irrealen Bedingungsgefüge, nur mittels eines Stimulus konsequent bei allen Sprechern evoziert werden. Beide Verfahren ergänzen sich somit gegenseitig und wirken bei der Analyse der Korpora der Gefahr entgegen, einzelne Artefakte falsch einzuschätzen.

Im folgenden werden nun die einzelnen, projektspezifischen Arbeitsschritte von der Erhebung bis hin zur konkreten Datenbankabfrage erläutert.

## **2. Die Arbeitsschritte**

### **2.1 Erhebung und Transkription von Dialektkorpora**

Die einzelnen Interviews werden mit einem Minidiskrecorder SONY MZ R35 aufgenommen. Anschließend erfolgt die Konvertierung des Lautmaterials in wav-Audiodateien. Dies ist für die digitale Transkription mit dem Programm PRAAT<sup>5</sup> unerlässlich, da nur Tonmaterial in diesem Datenformat als Audiofile eingelesen werden kann.

Zu diesem Zweck eignen sich kostenlose Konvertierungsprogramme, wie z.B. Wave-Recorder<sup>6</sup>. Der Wave-Recorder erlaubt exaktes Schneiden und Benennen der einzelnen Interviewteile. Alle so erstellten Audiofiles tragen eindeutige Siglen, so z.B. Bag1mDD1, d.h. der Sprecher stammt aus Bagnara Calabria, gehört der ersten Generation an, ist männlich und lebt in Deutschland. Die letzten beiden Stellen bezeichnen den Interviewteil D1, das spontansprachliche, größtenteils dialektale Gespräch.

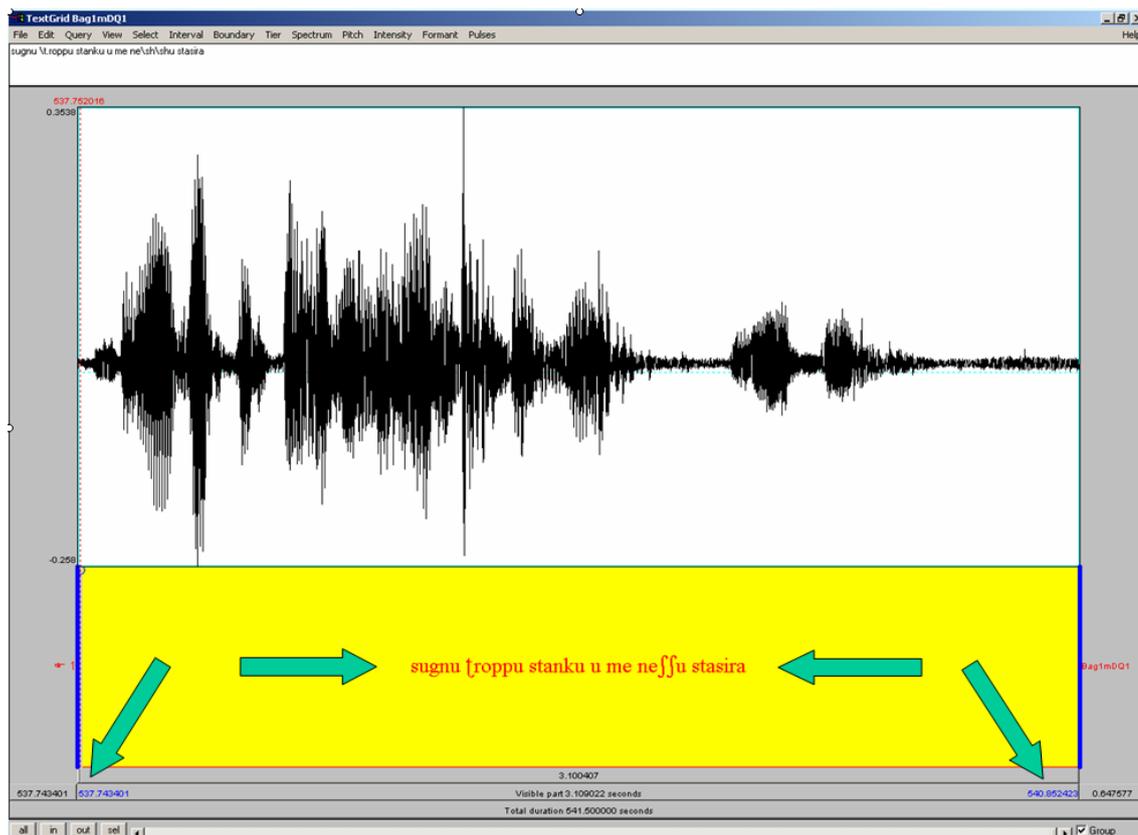
Die Transkription erfolgt mit dem Programm PRAAT. Dazu werden die Soundstreams bei der Transkription in Intervalle von ca. 2-8 Sekunden unterteilt.

---

<sup>5</sup>Genauere Informationen finden sich unter: <http://www.fon.hum.uva.nl/praat/>

<sup>6</sup>Die aktuellste Version ist downloadbar unter: <http://www.waverec.de/>

Abbildung Transkription in PRAAT



Idealtypische Intervalle entsprechen einem Turn, jedoch kann der Sprecher innerhalb eines Intervalls wechseln. In der Transkription wird dies durch Kürzel wie *E:* (*Explorer*) oder *I:* (*Informant*) angezeigt. Längere Pausen werden ebenfalls durch eigene Intervalle vom restlichen Tonmaterial abgetrennt, sodass zu einem späteren Zeitpunkt Datenbankabfragen des Typs “Wieviele Pausen gibt es im Gesamtkorpus?” oder etwa “Wie lange dauern die Pausen insgesamt?” gestellt werden können.

Insgesamt sind über 47 Stunden Tonaufnahmen transkribiert, wovon das Spontanmaterial ca. 33 Stunden umfasst. Dies entspricht einer Gesamtmenge von über 406.000 Token.

## 2.2 Von der Transkription zur Datenbank: Umwandlung von "Textgridfiles" in Tabellen

Angesichts der großen Menge an Sprachdaten und der zu bewältigenden quantitativen und qualitativen Analysen ist die Verwaltung nur in einer Datenbank sinnvoll und zeitökonomisch. Für die Verwendung in einer Datenbank müssen die transkribierten Texte jedoch zunächst in spezieller Weise umstrukturiert werden.

Den Ausgangspunkt dabei bilden die sog. "Textgridfiles". Es handelt sich um von PRAAT erzeugte, strukturierte Textdateien in ASCII-Kodierung. Sie enthalten, in

Listenform, zum einen den transkribierten Text, und zum anderen von PRAAT automatisch generierte Informationen über die einzelnen Intervalle, wie Intervallnummer und Anfangs- ("xmin") und Endpunkt ("xmax") des Intervalls.

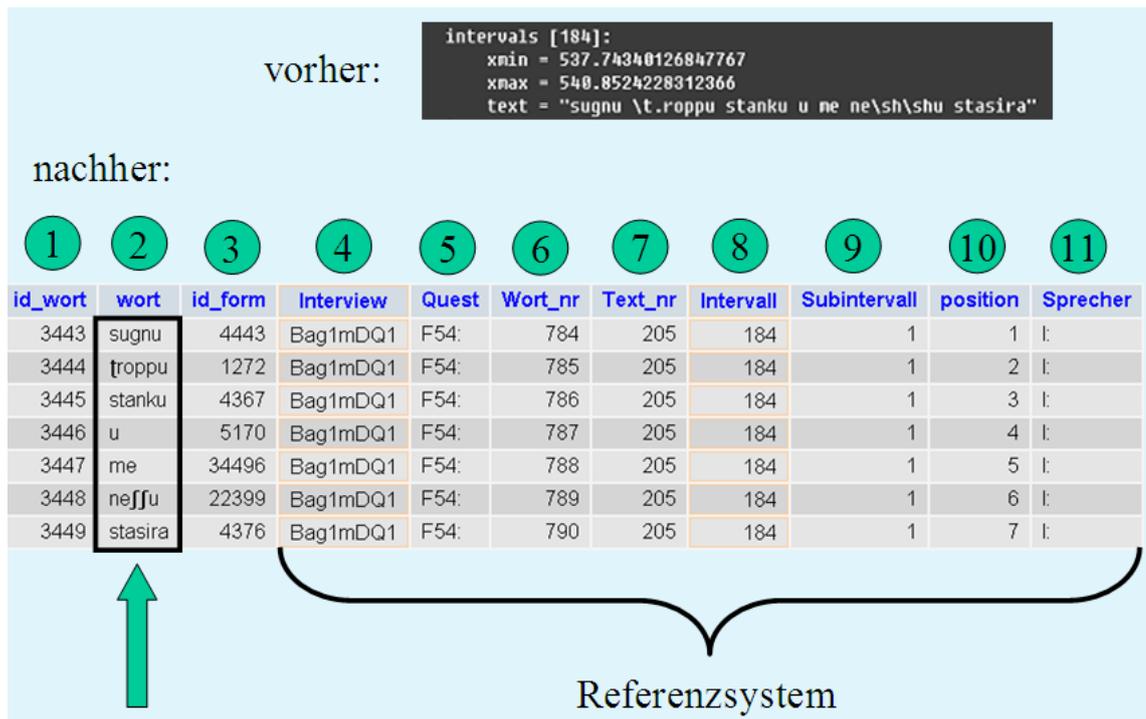
```
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 541.5
tiers? <exists>
size = 1
item []:
  item [1]:
    class = "IntervalTier"
    name = "BaglmDQ1"
    xmin = 0
    xmax = 541.5
    intervals: size = 185

[...]
intervals [54]:
  xmin = 162.58757242939296
  xmax = 165.18735770403987
  text = "si vvirgogna me nne\sh\shi râ gasa"
intervals [55]:
  xmin = 165.18735770403987
  xmax = 167.71738828298422
  text = "F17: "
intervals [56]:
  xmin = 167.71738828298422
  xmax = 170.34397546767903
  text = "mi piat\shi me vaju in dd\zhiru kê bbit\shi"
```

Für die Verwendung der Daten in einer Datenbank ist es nun nötig, die eindimensionale Listenstruktur der Textgridfiles in eine Tabellenstruktur zu überführen. Dieser Schritt erfolgt mit Hilfe kleiner Programme, die in unserem Fall in der aus der Unix-Welt stammenden Skriptsprache "awk" geschrieben wurden. Die Textgridfiles werden dabei dergestalt reorganisiert, daß zunächst insgesamt drei Tabellen entstehen: "Wort", "Text" und "Intervall". Die Tabelle "Wort" enthält pro Zeile je ein Wort des transkribierten Textes. Jedem Wort sind in weiteren Spalten verschiedene Metadaten wie z.B. die Nummer des Intervalls angegeben, in dem sich das jeweilige Wort befindet. Nach analogem Muster enthält die Tabelle "Text" pro Zeile jeweils den vollständigen transkribierten Text eines Intervalls. Die Tabelle "Intervall" nimmt pro Zeile im wesentlichen die Informationen über den Beginn und das Ende jedes einzelnen Intervalls auf.

Abbildung: Ausschnitt aus einem Textgridfile und das Resultat (Tabelle "Wort") nach der Bearbeitung durch das awk-Skript

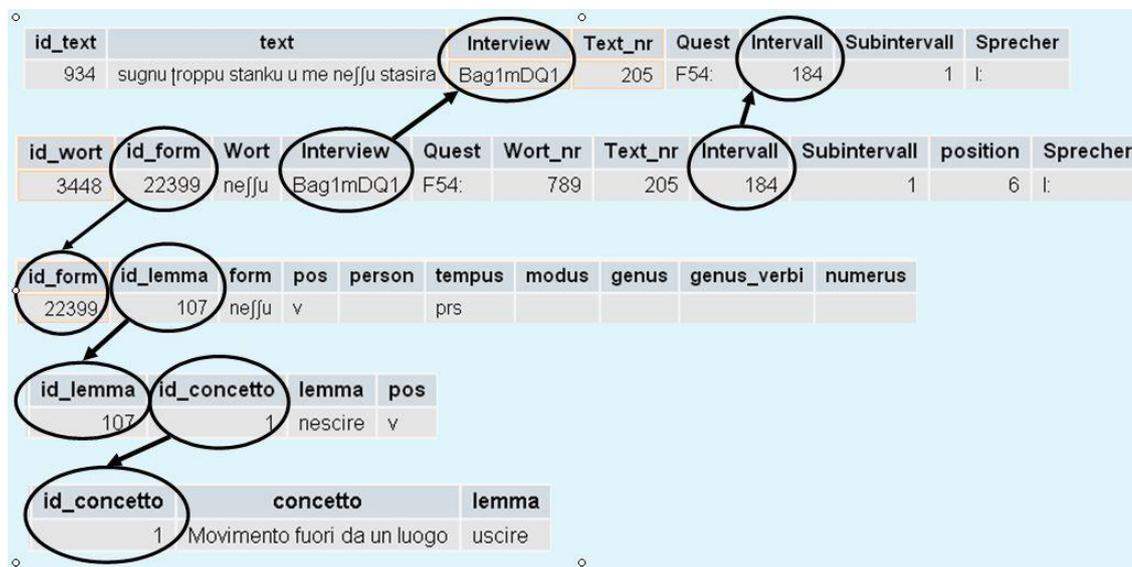


Vor dem Import der Tabellen in ein leistungsfähiges Datenbanksystem (MySQL) erfolgt schließlich noch die Umkodierung in UTF-8, einem weit verbreiteten Verfahren der Kodierung von Unicode-Zeichen, das die korrekte, plattformunabhängige Darstellung speziell der phonetischen Sonderzeichen garantiert.

### 2.3 Die Datenbanktabellen und ihre Funktionen

Die zentrale Einheit der Datenbank ist die Tabelle *wort*. Sie enthält alle Token und basiert auf den transkribierten Daten. Neben den aus den Transkriptionen (Textgridfiles) erzeugten Tabellen enthält die Datenbank eine Reihe weiterer Tabellen, die hauptsächlich dazu dienen, Redundanz von Daten zu vermeiden. Durch die Einführung einer eigenen Tabelle, die sog. "Types" enthält, ist es beispielsweise möglich, die Zuweisung von Attributen nur ein einziges Mal, nämlich in der Tabelle der Types, vorzunehmen und anschließend jedem Token nur noch die Schlüsselnummer (ID) des passenden Types zuzuordnen ("Tagging"). Dieses Verfahren ist zeit-ökonomisch und, was noch wichtiger ist, verringert die Gefahr von Fehleingaben. Nach demselben Prinzip sind auch die anderen Tabellen der Datenbank miteinander logisch verknüpft.

Die folgende Grafik verdeutlicht das Prinzip am Beispiel des Tokens *neffju* 'ich gehe (hin)aus' im Rahmen der Äußerung *sugnu troppu stanku u me neffju stasira* 'ich bin zu müde, um heute auszugehen'.



Umgekehrt ist durch die Formulierung geeigneter Abfragen das Auffinden zusammengehöriger Daten möglich. Dies ist vor allem im Zusammenhang mit Polymorphien, die in den Dialekten nicht die Ausnahme, sondern eher die Regel darstellen, von besonderem Interesse. Das verwendete System mehrerer Tabellen erlaubt es nämlich, dass beispielsweise bei der Suche nach dem standarditalienischen Negationspartikel *non* in der Tabelle der Lemmata alle sechs im Kalabresischen verwendeten unterschiedlichen Types (*non, nun, on, un, no, nu, u*) und die mit ihnen verbundenen Tokens gefunden werden.

In einigen Fällen ist es sinnvoll, eine dem Lemma übergeordnete Ebene, nämlich die des Konzepts (Tabelle "Concetto"), zu verwenden, wie z.B. im Fall von it. *dovere* und dial. *aviri* 'müssen'. Denn das dialektale Verb *aviri* kann neben der Grundbedeutung 'haben' in vielen Fällen die Funktion des standarditalienischen Modalverbs *dovere* annehmen.<sup>7</sup> Das Konzept OBLIGATION fasst somit unterschiedliche lexikalische Ausprägungen zusammen. Diese Ebene ermöglicht auch onomasiologisch orientierte Recherchen.

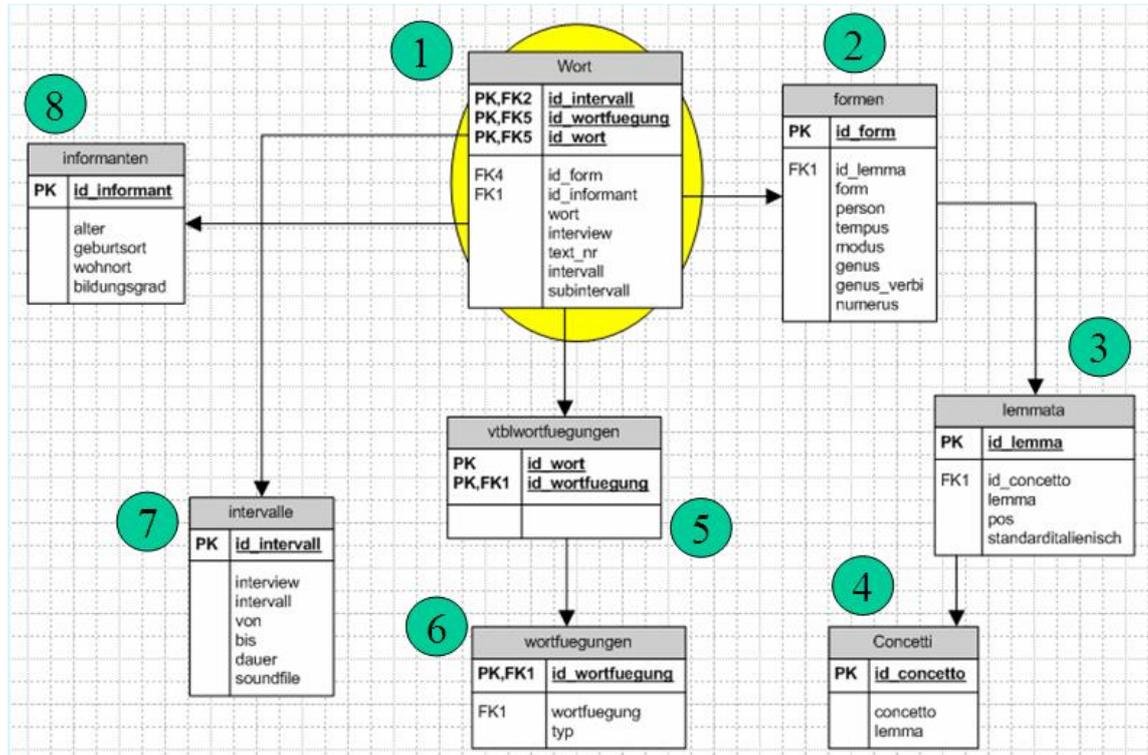
Die Tabellen *Wortfuegungen* und *vtblWortfuegungen* erlauben die Definition und Klassifizierung beliebiger und beliebig vieler Syntagmen innerhalb eines Intervalls.

<sup>7</sup> Im Dialektmaterial wird in 88% (!) der Fälle statt des italienischen *dovere* das panmeridionale *aviri* gebraucht. Vgl. Salming 2007, S.160

Dies ist in Intervallen, in denen mehrere Syntagmen (z.B. Nominal-/Verbalsyntagmen) analysiert werden, von großem Vorteil (s. unten Kap. 2.5).

Die folgende Grafik gibt einen Überblick über die in der Datenbank vorhandenen Tabellen und deren logische Verknüpfungen.

Abbildung: Die ASiCA -Datenbankstruktur (vereinfacht)



## 2.4 Das Tagging

Einer der Schwerpunkte der Projektarbeit liegt in der Zuordnung der rund 406.000 Token zu den passenden Types, dem sog. "Tagging". Da es sich nicht um ein standardsprachliches Korpus handelt, würde die automatische Zuweisung von Types zu Tokens zu einer enormen Fehlerquote führen. Hauptgrund ist das, für Nichtstandardvarietäten typische, zahlreiche Vorkommen von Homonymien. Allein das kalabresische Wort *u* repräsentiert sieben verschiedene Types (vgl. die folgende Tabelle).

Tabelle Okkurrenzen von dial. *u* im ASiCA -Korpus<sup>8</sup>

Form	Form-Nr.	Funktion	Lemma	Okkurrenzen
<i>u</i>	4779	Objektklitikon	<i>lo</i>	1346
<i>u</i>	5170	Konjunktion	<i>mu</i>	1458

<sup>8</sup> Tabelle aus Salminger 2007, S.81

u	5171	Definitiver Artikel	dial. <i>lu/</i> it. <i>il</i>	3709
u	5172	Negationspartikel	<i>non</i>	808
u	43985	Präposition+Artikel	<i>allo</i>	7
u	43983	Koordinierende Konjunktion	<i>o</i>	72
u	43986	Indefinitiver Artikel	<i>un</i>	7

Es ist also von Fall zu Fall zu entscheiden, welchem Type ein Token jeweils zuzuordnen ist. Zu diesem Zweck wurde in der Skriptsprache PHP eine spezielle Oberfläche programmiert, die dem Bearbeiter jeweils ein ausgewähltes Token im jeweiligen Kontext vorführt und gleichzeitig die Anzahl homonymer Types zur Auswahl anbietet. In schwierigen Fällen hat der Bearbeiter die Möglichkeit, sich die entsprechende Passage des entsprechenden Audiofiles vorspielen zu lassen (Lautsprechersymbol).

Abbildung: Tagging am Beispiel des Token *sugnu* (it. *sono* 'ich bin', 'sie sind')

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN ASiCA ONLINE Sito in allestimento! ATLANTE SINTATTICO DELLA CALABRIA

Exit Tagging Syntagmen Versionsinfo

Token Status POS geändert von

Token: *sugnu* Status: 1 POS: *geändert von: ISalminger*

SQL- select wort.id\_wort from wort where wort.sprecher not like 'E%' and wort.wort regexp 'sugnu' and wort.checked regexp '1' and  
 Abfrage: wort.changed\_by like 'ISalminger'

Treffer 1 von 42 < > Gehe zu Treffer:

Wort: *sugnu* - ID: 3443 - Interview: Bag1mDQ1 - Text-Nr. 205 - Sprecher: I:

. / # vo anguna gosa te fattsu nu kaffe o nu lattte mmandorle o u mpestat i domande magari kiju n si virgogna e mi ritfj si / Hai già mangiato qualcosa ? / mandžasti nguna gosa ? / Sono troppo stanco/a per uscire stasera . /

*sugnu* troppo stanku u me neffu stasira

Token Intervall anhören

	form	lemma	pos_form	person	tempus	modus	genus	genus_verbi	numerus	id_form
<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	<i>dubbio</i> ✓	<i>sugnu</i>	<i>essere</i>	<i>V</i>	<i>1</i>	<i>prs</i>	<i>ind</i>		<i>sg</i>	<i>4443</i>
<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	<i>dubbio</i> ✓	<i>sugnu</i>	<i>essere</i>	<i>V</i>	<i>3</i>	<i>prs</i>	<i>ind</i>		<i>pl</i>	<i>30373</i>

Types

Neue Form anlegen - Neues Lemma anlegen

Die Möglichkeit, einem bestimmten Token automatisch einen Wert zuzuweisen, ist ebenfalls gegeben (grünes Häkchen vor dem Type), jedoch nur in wirklich eindeutigen Fällen empfehlenswert. Die Funktionen *Neue Form anlegen* bzw. *Neues Lemma anlegen* erleichtern die kontinuierliche Datenbankpflege.

Insgesamt müssen schätzungsweise 60% des Gesamtmaterials manuell getaggt werden, was einen beachtlichen Zeitaufwand bedeutet. Derzeit sind etwa 130.000 Token eindeutig identifiziert. Je vollständiger das Tagging, desto verlässlicher sind die Abfrageergebnisse.

#### 4. Die Abfragen

Das Auffinden der oft komplexen Syntaxkonstruktionen erfolgt mittels sog. "SQL"-Abfragen, welche alle Vorkommnisse einer speziellen morphologischen oder morphosyntaktischen Struktur herausfiltern.

Während die Ergebnisse des Fragebogenmaterials noch mittels einer einfachen Abfrage innerhalb der Tabellen *Text* oder *Wort* gefunden werden können<sup>9</sup>, sind vor allem bei der Suche nach bestimmten Mustern im Spontanmaterial komplexe Abfragen über mehrere Tabellen notwendig.

Wenn beispielsweise gezielt nach dem Serialisierungsverhalten des Modalverbs *volere* 'wollen'<sup>10</sup> gesucht werden soll, so muss die Abfrage über vier Tabellen gestellt werden.

SQL-Beispiel: JOIN-Abfrage über 4 Tabellen

```
SELECT wort_1.Wort AS Wort1, lemma_1.lemma AS lemma1, wort_2.Wort AS
Wort2, wort_3.Wort AS Wort3, text.Interview, text.Text_nr, text.Text
FROM text
JOIN wort AS wort_1 ON text.Interview = wort_1.Interview AND
text.Text_nr = wort_1.Text_nr
JOIN formen AS form_1 ON wort_1.id_form = form_1.id_form
JOIN lemmata AS lemma_1 ON form_1.id_lemma = lemma_1.id
JOIN wort AS wort_2 ON text.Interview = wort_2.Interview AND
text.Text_nr = wort_2.Text_nr
JOIN formen AS form_2 ON wort_2.id_form = form_2.id_form
JOIN wort AS wort_3 ON text.Interview = wort_3.Interview AND
text.Text_nr = wort_3.Text_nr
JOIN formen AS form_3 ON wort_3.id_form = form_3.id_form
WHERE lemma_1.lemma = 'volere'
AND (form_2.pos LIKE '%C1%')
OR form_2.pos LIKE '%CP%')
AND (form_3.tempus LIKE '%prs%'
```

---

<sup>9</sup> Der Fragekatalog enthält 54 italienische Stimuli, welche in der Transkription z.B. durch *F12* für *frase 12* gekennzeichnet sind. Will man die Ergebnisse zum Stimulus *F12: Se avessi più tempo, lo farei immediatamente* 'Wenn ich mehr Zeit hätte, würde ich es sofort tun' angezeigt bekommen, so genügt eine Abfrage mit einer einfachen WHERE-Bedingung, "SELECT \* FROM `text` WHERE `Quest` = 'F12:'"

<sup>10</sup> In diesem Zusammenhang sei erwähnt, dass eine der augenfälligen Besonderheiten der kalabrischen Syntax der eingeschränkte Gebrauch des Infinitivs ist (vgl. dazu etwa Rohlfs 1949: §717 u. Rohlfs 1985). Dies gilt insbesondere nach dem Modalverb *volere*, das bei Subjektkoreferenz mit einer Konjunktion und konjugiertem Verb konstruiert und gewissermaßen prototypischer Auslöser für die infinitivlose Komplementierung zu sein scheint (vgl. Rohlfs 1949: §717). Allerdings relativieren neuere Untersuchungen zur Verbreitung des Infinitivs in südkalabrischen Varietäten diese strikte Ansicht (vgl. Salminger 2007).

```

OR form_3.tempus LIKE '%impf%')
AND text.Interview LIKE '%D1'
AND (wort_3.id_wort - wort_1.id_wort <=2
AND wort_3.id_wort - wort_1.id_wort >0)
AND (wort_2.id_wort - wort_1.id_wort <=1
AND wort_2.id_wort - wort_1.id_wort >0)
AND text.sprecher NOT LIKE 'E%'

```

Mit dieser SQL-Abfrage werden diejenigen Kontexte gefunden, in denen ein konjugiertes Verb im Präsens oder Imperfekt einer Konjunktion folgt, welche gleichzeitig von einer Verbform abhängt, die dem Lemma *volere* zugehörig ist. Zugleich gilt die Bedingung, dass zwischen Konjunktion und abhängigem Verb nichts stehen darf und zwischen Matrixverb und abhängigem Verb maximal eine Stelle treten darf. Eine weitere Einschränkung ist, dass ausschließlich im Spontanmaterial gesucht wird und der Sprecher vom Explorator verschieden sein muß.

Auf diese Weise finden sich Belege wie

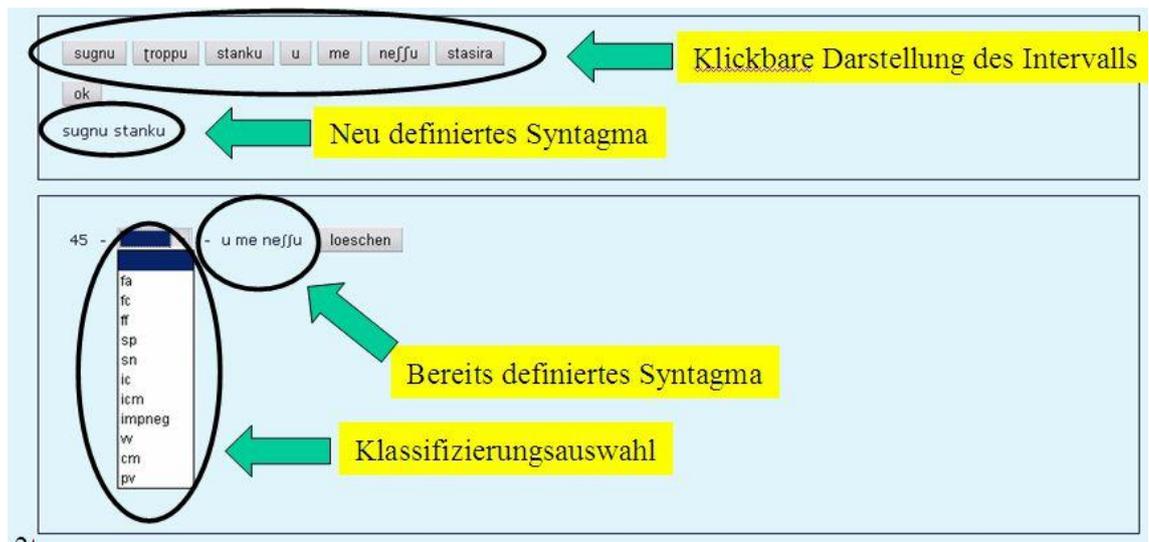
- (1) *tu vo u rimani kkha ?*  
 Du willst, dass (du) bleibst hier ?  
 'willst du hier bleiben'

Die Ergebnisliste zeigt derzeit 105 Treffer an, die dieser syntaktischen Anordnung folgen<sup>11</sup>.

Mit derartigen Abfragen kann bereits ein Großteil der relevanten Syntaxphänomene herausgefiltert werden. Dennoch ist in einigen Fällen von einer Automatisierung des Syntagmentagging dringend abzuraten, da in der Spontansprache oft unvorhersehbare Abweichungen (z.B. Satzabbrüche) vom erwarteten Syntaxmuster auftreten. Aus diesem Grund muß die durch die SQL-Abfrage gewonnene Trefferliste auf Wortfügungsebene weiter verfeinert und bezüglich ihrer Syntagmen näher spezifiziert werden. Diesem Zweck dient die bereits weiter oben (s. Kap. 2.3) erwähnte Möglichkeit, innerhalb eines Intervalls mehrere Tokens zu Syntagmen zu verbinden und diese zu klassifizieren. Die Definition der Syntagmen erfolgt dabei über eine eigens dafür programmierte Oberfläche:

---

<sup>11</sup> Im Gegensatz dazu finden sich 261 Okkurrenzen, in denen *volere* im Fall koreferenter Subjekte ein infinitivisches Verbalkomplement aufweist. Die Infinitivkonstruktionen sind diatopisch entweder als nordkalabresisch (Luzzi, Cariati Belvedere) oder als typisch für die Dialekte des äußersten Südens (Rosarno, Oppido, Bagnara) einzureihen. Die mittelkalabresischen Dialekte mit den Punkten San Pietro, Amaroni, Bivongi, Polistena ziehen nach *volere* eher die infinitivlose Variante vor.



## 2.5 Von der Datenbank zum Online-Atlas

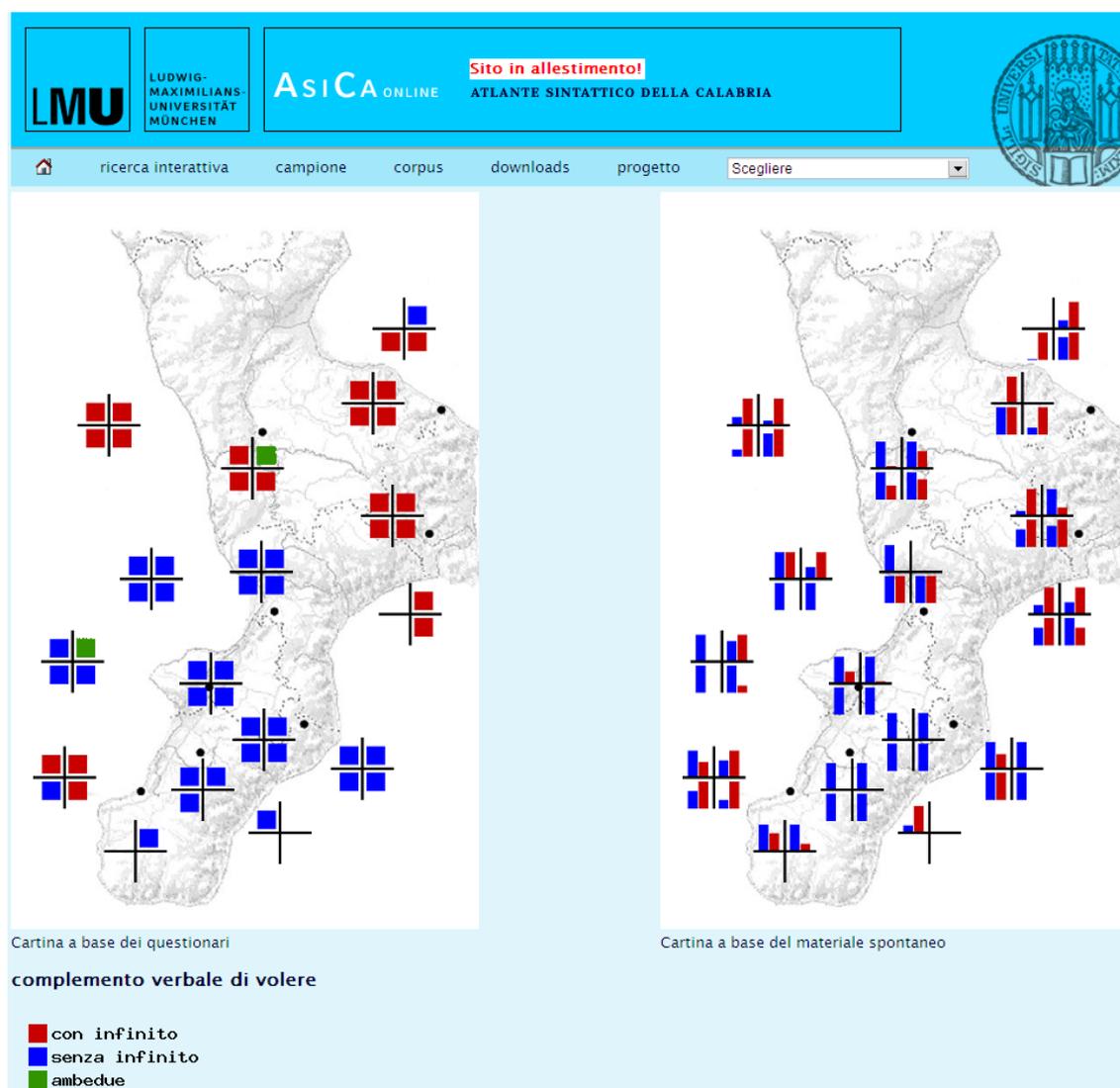
Die bisher vorgestellten Funktionen stehen alle im Zusammenhang mit der Dateneingabe und -pflege, anders gesagt: der Administration der Datenbank. Eigentliches Ziel des Projektes ist jedoch, das gesammelte Material und dessen Analyse interaktiv der Öffentlichkeit zugänglich zu machen. Im Focus steht dabei die Visualisierung von Analyseergebnissen in Form von interaktiven Karten.

Der öffentliche Bereich von Asica ist zu erreichen über folgenden Link: <http://asica.gwi.uni-muenchen.de>. Derzeit werden dort folgende Optionen angeboten: Der Link *Campione* verweist auf Interviewausschnitte von ethnographischem Interesse. Links ist das Transkriptionsoriginal zu sehen, rechts die Übersetzung ins Standarditalienische. Am äußersten linken Rand befindet sich eine Toolbar, mit der das Audiofile abgespielt wird.

Unter *Corpus* sind sämtliche Transkriptionen sowie das jeweilige Audiomaterial wählbar.

Das Herzstück von Asica, die Atlas-Funktion, ist über den Link *Ricerca interattiva* zu erreichen. Hier lassen sich über ein Drop-Down-Menü verschiedene syntaktische Analysen anwählen, deren Ergebnisse sodann auf zwei Karten von Kalabrien visualisiert werden. Dabei stellt die linke Karte die Ergebnisse des Fragebogens (qualitative Analyse) und die rechte jene des Spontanmaterials (quantitative Analyse) dar. Form und Farbe der Diagrammsymbole werden stets aktuell auf Basis der aus der Datenbank mittels passender SQL-Abfragen gewonnen Werte errechnet (die Arbeit erledigt ein PHP-Skript). Momentan ist die Auswahl auf zwei Musterfälle beschränkt. Eine hohe Zuverlässigkeit der dargestellten Ergebnisse ist überdies erst nach dem vollständigen Abschluß des Taggings gewährleistet.

Abbildung: Recherchebeispiel "Komplement des Verbs *volere* ('wollen')"



### 3. Perspektiven des ASICA

Der online-Atlas des ASICA sowie die zugrundeliegende Datenbank werden derzeit noch optimiert. In Zukunft können sämtliche Ergebnisse aus den beiden Interviewteilen visualisiert werden. Geplant ist ferner die Entwicklung einer Funktion, die den teilweisen oder auch vollständigen Export des transkribierten Korpus im XML-konformen<sup>12</sup> CES-Standard<sup>13</sup> erlaubt.

Aktuelle Arbeiten zeigen (Krefeld 2007b, Salminger 2007, Schwägerl 2007), dass die Feststellung, die lautliche Einheit einer Dorfmundart sei ein Mythos (Jaberg/Jud 1928: 216), auch für andere sprachliche Teilbereiche, wie Lexik und Morphosyntax, zutrifft.

<sup>12</sup> "Extensible Markup Language"; vgl. [http://de.wikipedia.org/wiki/Extensible\\_Markup\\_Language](http://de.wikipedia.org/wiki/Extensible_Markup_Language)

<sup>13</sup> "Corpus Encoding Standard"; vgl. <http://www.cs.vassar.edu/CES/>

## ASiCA Zugangsdaten

<https://pma.gwi.uni-muenchen.de> (Benutzerkennung: ascstud; Paßwort: Bivongi)

## Bibliographie

Jaberg, Karl/ Jud Jakob 1928-40: *Sprach- und Sachatlas Italiens und der Südschweiz*. Zofingen: Ringier

Jaberg, Karl/ Jud, Jakob 1928: *Der Sprachatlas als Forschungsinstrument. Kritische Grundlegung und Einführung in den Sprach- und Sachatlas Italiens und der Südschweiz*. Halle/Saale: Niemeyer

Krefeld, Thomas 2005: Dialeto, variazione linguistica - e l'AIS. Come l'ideologia ha (de)formato la modellazione dello spazio linguistico; in: *Romanistisches Jahrbuch* 55, 83-103

Krefeld, Thomas 2006 [forthcoming]: Räumlich fundierte Varietätenlinguistik - eine Option und ihre Geschichte; in: Bernhard, Gerhard / Siller, Heidi (Hrsg.): *Orientierungen im Raum*.

Krefeld, Thomas 2007a [forthcoming]: Dal punto diatopico alla diatopia del punto: Una prospettiva promettente - ma trasformativa; in: Raimondi, Giammaria (ed.): *Dove va la dialettologia? La dialectologie d'aujourd'hui. Actes du colloque international. Saint-Vincent-Aoste-Cogne, 21-23 septembre 2006*

Krefeld, Thomas 2007b [forthcoming]: Kalabresisch: *fra pogu vegnu a ti*  $\nabla$ rovu. Fossil oder Produkt rezenter syntaktischer Wandels; in: Schmidt-Riese, Roland/Stark, Elisabeth/Stoll, Eva (Hrsg.) 2007: *Romanische Syntax im Wandel*. Tübingen: Narr

Radtke, Edgar 1988: Italienisch Areallinguistik IX: Kampanien. Kalabrien; in: Holtus, Günter/ Metzeltin, Michael/ Schmitt, Christian (eds.): *Lexikon der romanischen Linguistik*. Band IV: *Italienisch. Korsisch. Sardisch*. Tübingen: Niemeyer

Rohlf, Gerhard 1949: *Historische Grammatik der italienischen Sprache und ihrer Mundarten*, Band II: *Syntax und Formenlehre*. Bern: Francke

Rohlf, Gerhard 1985: *Latinità ed ellenismo nel Mezzogiorno d'Italia. Studi e ricerche. Dalla Magna Grecia alla Grecia italiana*. Chiaravalle: Framasud

Ruffino, Giovanni 1995 (eds.): *Percorsi di geografia linguistica. Idee per un atlante siciliano della cultura dialettale e dell'italiano regionale*. Palermo: Centro di studi filologici e linguistici siciliani

Salminger, Irmengard 2002: Variazione sintattica in contesto migratorio: L'uso dell'infinito nel calabrese di Cutro; in: Krefeld, Thomas (ed.): *Spazio vissuto e*

- dinamica linguistica. Varietà meridionali in Italia e in situazione di extraterritorialità.*  
Frankfurt: Lang, 149-168
- Salminger, Irmengard 2007: *Kalabrische Junktionstechniken zwischen Aggregation und Integration gezeigt am ASICA*. München (= bislang unveröffentlichte Dissertation; Universität München)
- Schwägerl, Verena 2007: Die Realisierung der Bedingungssätze in den Dialekten Kalabriens (= unveröffentlichte Seminararbeit; Universität München); <http://asica.gwi.uni-muenchen.de/downloads/VerenaSchwaegerlperiodiipoteticiASICA.pdf>
- Trumper, John 1997: Calabria and Southern Basilicata; in: Maiden, Martin/Mair, Parry (eds.): *The dialects of Italy*. London: Routledge, 355-364
- Vincent, Nigel 1997: Complementation; in: Maiden, Martin/Mair, Parry (eds.): *The dialects of Italy*. London: Routledge, 171-178