

## Die Rolle der Phonologie in der multilingualen Sprachtechnologie\*

### 1 Einführung

Multilinguale Kommunikation fordert die Übertragbarkeit der gängigen Techniken auf eine große Anzahl verschiedener Sprachen und stellt damit hohe Anforderungen an die Sprachtechnologie. Obwohl die Spracherkennung und die Sprachsynthese sich in den letzten Jahren in vielen Bereichen unseres Lebens durchgesetzt haben, sind die Anwendungen häufig von einer bestimmten Domäne abhängig. Auskunftssysteme, zum Beispiel, sind so beschaffen, dass Probleme aufkommen, sobald neue Wörter erkannt oder produziert werden sollen. Während die größeren Sprachen der Welt von jeder neuen Entwicklung im sprachtechnologischen Bereich profitieren, leiden die kleineren Sprachen eher unter einem Mangel an Ressourcen und Anwendungen. Dabei ist besonders für bedrohte Sprachen der Aufbau linguistischer Ressourcen, die bei der Spracherkennung oder Sprachsynthese eingesetzt werden können, von enormer Wichtigkeit.

Die meisten kommerziellen Spracherkennungssysteme bauen auf stochastischen Methoden auf, die zwar implizit linguistische Informationen beinhalten, jedoch nicht in einer Form, die sich zum Beispiel unmittelbar für die Sprachsynthese eignet. Das linguistische Wissen liegt im System versteckt und kann nicht ohne weiteres explizit aufgedeckt und weiterverwendet werden. Jedes neue System bedarf einer erneuten „Trainingsphase“, damit zusätzliches Wissen hinzugefügt werden kann. Dies setzt einen großen Datensatz voraus, der auch entsprechend annotiert werden muss. Sinnvoller scheint es hingegen, alle Daten, die für neue Anwendungen gesammelt werden, so zu repräsentieren, dass sie von verschiedenartigen Technologien weiterverwendet werden können.

Der Schwerpunkt bei der weiteren Entwicklung in der Sprachtechnologie muss auf generische Techniken gelegt werden, die für alle Sprachen gelten. Nur so kann eine robuste multilinguale Sprachtechnologie entstehen, die auch skalierbar ist. Um die Portabilität sprachtechnologischer Anwendungen mit Blick auf „neue“ Sprachen zu gewährleisten, ist die Entwicklung innovativer Methoden für die Akquisition und Repräsentation linguistischer Informationen verschiedener Granularitäten unumgänglich. Mit Blick auf diesen Aspekt wird in diesem Papier eine neuartige Modellierung phonologischer Informationen präsentiert, die sowohl in dem hier vertretenen compu-

---

\* Dieser Beitrag geht auf Arbeiten zurück, die von der Science Foundation Ireland unter Grant Nr. 02/IN1/I100 gefördert werden. Die hier vertretenen Meinungen, Ansichten, Schlussfolgerungen und Empfehlungen sind diejenigen der Autoren, und geben nicht unbedingt die Ansichten der Science Foundation Ireland wieder.

terlinguistischen Ansatz zur Sprachtechnologie Anwendung findet, als auch in traditionellen Systemen integrierbar ist.

Um den Rahmen für die Diskussion zu legen, werden im Folgenden die relevanten Aspekte des *Time Map* Modells (Carson-Berndsen, 1998, 2000) vorgestellt, die als Grundlage für die Entwicklung portabler Modelle dienen. Hierfür ist zunächst in Abschnitt 2 die Einführung der phonotaktischen Automaten und der multilinearen Repräsentationen notwendig. Danach wird in Abschnitt 3 gezeigt, wie dieses Modell in der Spracherkennung eingesetzt wird. In Abschnitt 4 wird das Konzept der *multilingual time map* erläutert, auf welchem das Lexikon basiert. Die Kodierung dieses lexikalischen Wissens, sowie formale Aspekte der Umwandlung in ein portables Transferformat werden in Abschnitt 5 diskutiert; Abschnitt 6 stellt am Beispiel einer konkreten *multilingual time map* deren Struktur und diverse benutzerorientierte Ausgabeformate dar.

## 2 Phonotaktische Automaten und multilineare Repräsentationen

Als Grundlage für das *Time Map* Modell dient der so genannte phonotaktische Automat, eine Netzwerkrepräsentation aller zulässigen Lautkombinationen einer Sprache. Ein Ausschnitt eines solchen phonotaktischen Automaten, der die CC-Kombinationen im deutschen Silbenanlaut beschreibt, befindet sich in Abbildung 1.

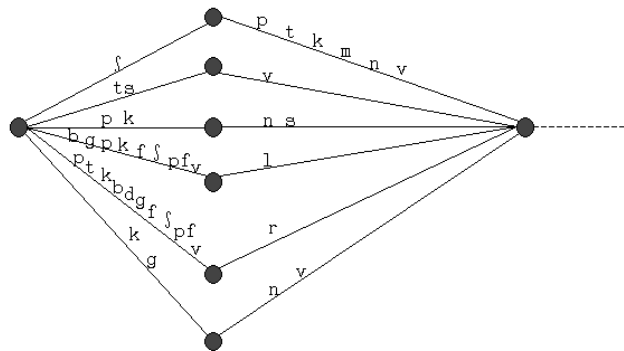


Abb. 1: CC-Lautkombinationen im deutschen Silbenanlaut.

Aufgrund der Tatsache, dass alle zulässigen Lautkombinationen der Sprache als phonotaktischer Automat erfasst sind, werden durch diesen Repräsentationsformalismus

die Intuitionen des Muttersprachlers in Bezug auf phonologische Wohlgeformtheit modelliert. Obwohl die Form *Blied* ([βλi|τ]) im Deutschen nicht existiert, kann jeder Muttersprachler sie als mögliches deutsches Wort dadurch erkennen, dass diese Form die phonotaktischen Einschränkungen des Deutschen widerspiegelt. Im Gegensatz dazu, ist die Kombination [Tp] in deutschen Silbenanlaut nicht zulässig: sie kommt jedoch im englischen Silbenanlaut vor (siehe Abbildung 2) - wie im Wort *throw* ‚werfen‘. Gleichermäßen die Kombination [kv] wie *Knie* ist im Deutschen zulässig, die englische Graphemkombination <kn> wie in *knee* wird jedoch als [v] ausgesprochen – die Kombination [kv] ist im Englischen nicht wohlgeformt.

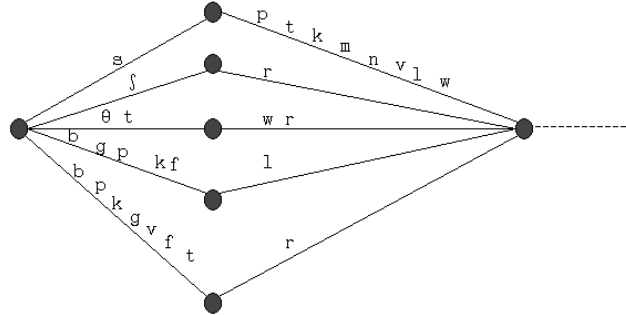


Abb. 2: CC-Lautkombinationen im englischen Silbenanlaut.

Der phonotaktische Automat erlaubt es, zwischen wohlgeformten und nicht-wohlgeformten Silben (systematischen Lücken) einer Sprache zu unterscheiden. Außerdem wird ein Lexikon benötigt, um die lexikalisierten von den potenziellen Formen (zufälligen Lücken) zu unterscheiden. Somit ist die Phonotaktik sprachspezifisch – es kann also für jede Sprache ein spezieller phonotaktischer Automat entwickelt werden (mehr dazu in Abschnitt 4).

Der phonotaktische Automat wird im *Time Map* Modell eingesetzt, um multilineare Repräsentationen von Sprachäußerungen phonologisch zu interpretieren. Eine Sprachäußerung wird als autosegmentale Merkmalsrepräsentation (vgl. Goldsmith 1990) dargestellt, in der phonetische und phonologische Merkmale in zeitlich überlappenden Relationen und in Präzedenzrelationen zueinander stehen können. Es wird also davon abgesehen, das Sprachsignal in streng nicht-überlappende Segmente zu teilen. Stattdessen wird eine Ereignislogik benutzt, Koartikulationsphänomene zu modellieren. Eine multilineare Repräsentation des deutschen Wortes *Knie* befindet sich in Abbildung 3. Hieraus wird deutlich, dass die Merkmale asynchron sind, d.h. sie haben unterschiedliche Start- und Endpunkte.

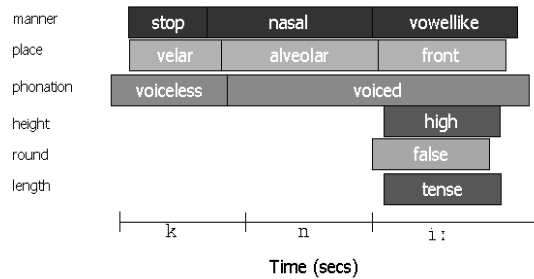


Abb. 3: Multilineare Repräsentation des Wortes *Knie*.

Um den phonotaktischen Automaten zum Zweck der Interpretation einer solchen Repräsentation anzuwenden, müssen die atomaren Symbole, die an den Übergängen stehen, durch Einschränkungen erweitert werden, die die Überlappungsrelationen für die jeweiligen Phoneme beschreiben. Anstatt eines  $[\Sigma]$  müssen also drei Beschränkungen erfüllt werden, die die Überlappungsrelation ( $\circ$ ) zwischen den phonologischen Merkmalen für diesen Laut beschreiben:

$$\begin{aligned} C_1 &: \text{fricative} \circ \text{voiceless} \\ C_2 &: \text{fricative} \circ \text{palato-alveolar} \\ C_3 &: \text{palato-alveolar} \circ \text{voiceless} \end{aligned}$$

Mit dem Einsatz des *Time Map* Modells in der Spracherkennung stellt sich die Frage, wie eine multilineare Repräsentation zustande kommt. Diese wird im folgenden Abschnitt erläutert.

### 3 Das Time Map Modell in der Spracherkennung

Im Gegensatz zu traditionellen Techniken in der Spracherkennung, baut der *Time Map* Ansatz auf sprachspezifischen phonologischen Strukturen auf, die das vollständige Silbenvokabular der Sprache aufweisen. Ein Vorteil dieses Ansatzes ist, dass der phonotaktische Automat einfach ausgetauscht werden kann, wenn eine andere Sprache erkannt werden soll. Um eine komplette Portabilität im Sinne der Multilingualität zu gewährleisten, ist jedoch die Erweiterung der Merkmalsmenge notwendig. Letzteres wird im Zusammenhang mit dem Lexikon definiert.

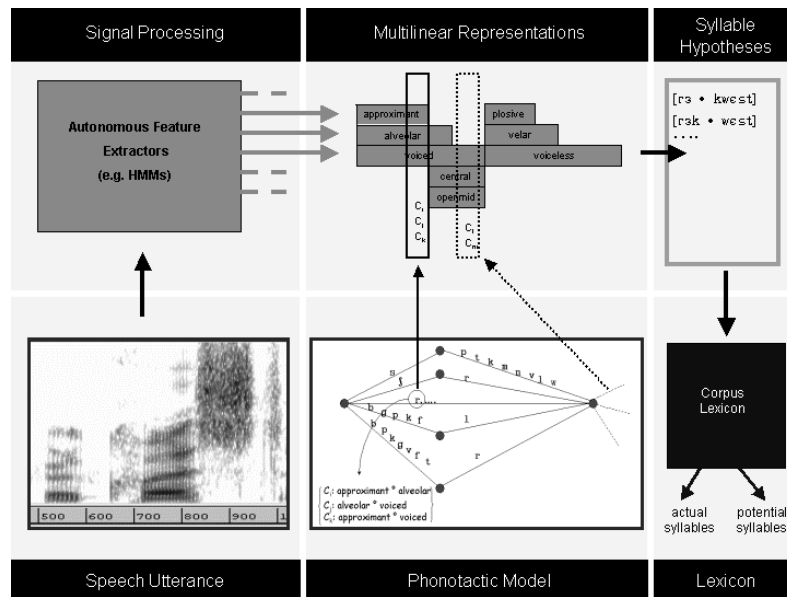


Abb. 4: Spracherkennungs-Architektur des *Time Map* Modells.

Die Spracherkennungs-Architektur des *Time Map* Modells wird in Abbildung 4 dargestellt. Aus dem Sprachsignal werden Merkmale extrahiert, die als multilineare Repräsentationen im Sinne von Abschnitt 2 dargestellt werden. Hierfür ist eine spezielle Technik entwickelt worden, die auf parallele *hidden Markov* Modelle aufbaut (siehe Abu-Amer/Carson-Berndsen, 2003). Der phonotaktische Automat definiert Beschränkungen, die in der multilinearen Repräsentation zu erwarten sind. Wenn diese erfüllt sind, werden in dem nächsten Zeitintervall die Einschränkungen für den nächsten Übergang gesucht, bis der phonotaktische Automat einen Endzustand erreicht hat. In diesem Fall ist eine wohlgeformte Silbe gefunden worden. Sollten die Einschränkungen nicht erfüllt werden, werden entweder die Bedingungen gelockert, oder die multilineare Repräsentation wird als nicht wohlgeformt verworfen (mehr hierzu in Carson-Berndsen & Walsh in press). Die so gewonnenen Silbenhypothesen werden dann an das Lexikon geschickt, wo entschieden wird, ob diese Formen lexikalisiert oder nur potenziell sind. Letztere Formen erhalten einen niedrigeren Konfidenzwert.

Eine Umkehrung dieser Architektur für den Einsatz in der Sprachsynthese wird zur Zeit erforscht (vgl. Bohan et al. 2001). In diesem Fall werden multilineare Repräsentationen anhand von den - in dem phonotaktischen Automaten definierten - Überlappungs- und Präzedenzrelationen erzeugt und die Abbildung zwischen Merkmalen und Syntheseparametern gelernt. Ziel dieser Untersuchungen ist es, ein wohldefiniertes,

portables Format für die Beschreibung von phonologischen Informationen zu finden, die eine multilinguale Funktionalität der Sprachtechnologie ermöglicht. Im nächsten Abschnitt wird eine Erweiterung des phonotaktischen Automaten vorgestellt, die genau diesem Zweck dient.

#### 4 Multilingual Time Maps

Um phonetische und phonologische Informationen verschiedener Granularität zu integrieren, ist das Konzept des phonotaktischen Automaten erweitert worden. Daraus entstanden ist eine *multilingual time map* (siehe Carson-Berndsen, 2002). Sie unterscheidet sich von dem phonotaktischen Automaten dadurch, dass die Übergänge nicht nur Phoneme und Überlappungsrelationen aufweisen, sondern auch Grapheme, Allophone, Merkmale, Durchschnittslänge, Häufigkeit und Wahrscheinlichkeit. Ein Übergang aus der *multilingual time map* für deutsche CC-Lautkombinationen (analog zu Abbildung 1) wird in Abbildung 5 dargestellt.

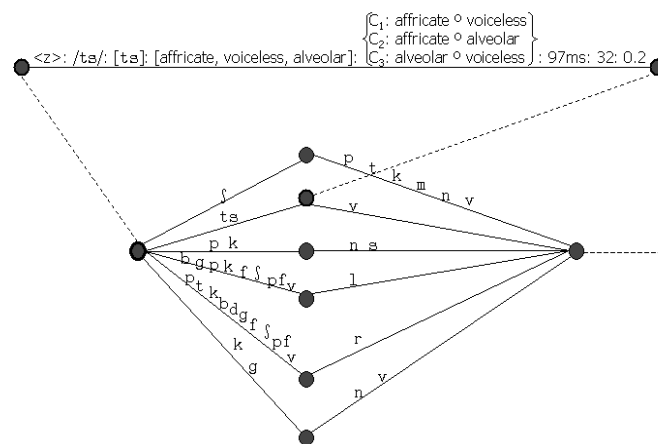


Abb. 5: Ein Übergang in der *multilingual time map*.

Eine *multilingual time map* kann als mehrbändiger Automat interpretiert werden und kann außerdem zum Zweck der Graphem-Phonem-Übersetzung eingesetzt werden. Dadurch, dass eine *multilingual time map* auf endlichen Automaten basiert, existieren anerkannte Lern-Algorithmen, die zum Zweck des phonotaktischen Lernens eingesetzt

werden können (siehe Carson-Berndsen, Gut & Kelly, 2003). Ausgehend von einem annotierten Sprachkorpus werden Partitionen im Sinne der Silbenstruktur der jeweiligen Sprache erstellt (z.B. CCC-Kombinationen, CC-Kombinationen, VC-Kombinationen usw). Für jede Partition wird ein phonotaktischer Automat gelernt. Die Partitionen werden dann zusammengefügt, um eine für den Korpus vollständigen Phonotaktik aufzubauen. Um die zufälligen Lücken zu füllen, die auf Grund der Tatsache, dass ein Korpus nicht alle Formen abdecken kann, zwangsweise entstehen, wird eine Kombination von Techniken eingesetzt, die sowohl auf natürlichen Merkmalsklassen als auch auf visuellen Informationen basieren. Eine Bestätigung der hypothesierten Form wird durch einen Muttersprachler vorgenommen und die Lücken werden gefüllt. Diese Vorgehensweise wird in Abbildung 6 verdeutlicht.

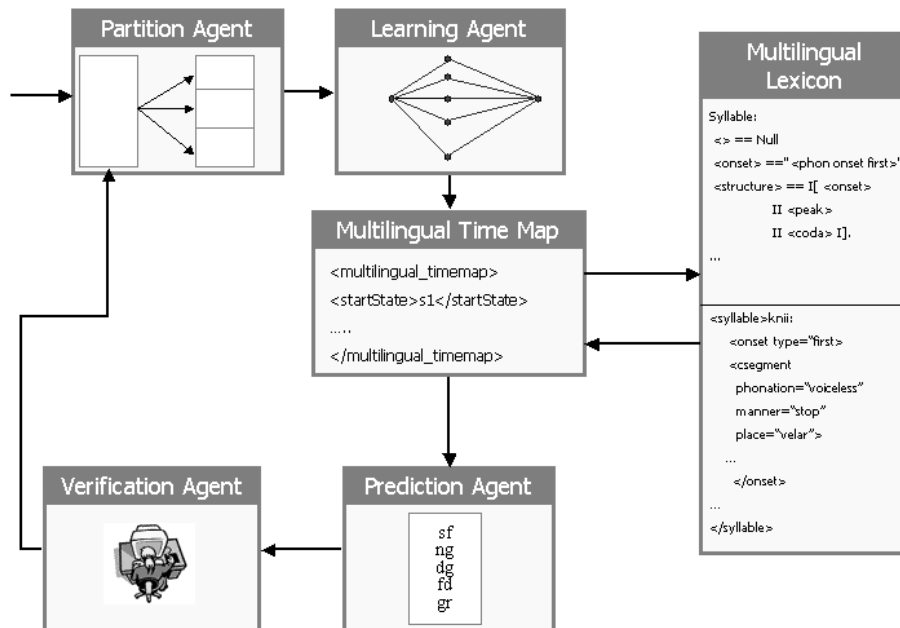


Abb. 6: Lernen einer *multilingual time map*.

## 5 Die Struktur lexikalischer phonologischer Repräsentationen

Ausgehend von dem Konzept der *multilingual time map* und seiner Anwendung sowohl zum Zweck der Sprachsynthese, sowie der Spracherkennung stellt sich die Frage nach der Portabilität der lexikalischen Information für weitere Anwendungen. Das Hauptinteresse in bisherigen Ansätzen zur Architektur multilingualer Lexika liegt in der Lexikonentwicklung (vgl. Cahill/Gazdar 1999) oder darin, Gemeinsamkeiten zwischen verwandten Sprachen mittels hierarchischer Vererbungsmechanismen zu erfassen (vgl. Tiberius 2001). Im Gegensatz dazu ist das Ziel der *multilingual time maps* die Bereitstellung generischer, wieder verwendbarer Anwendungen, die die Entwicklung und Evaluation phonologischer Systeme ermöglichen.

Um nun ein Lexikonformat bereitzustellen, das phonologische Information nicht nur für Synthese und Erkennung, sondern für beliebige Anwendungen bereitstellt, wird im Folgenden ein portables, generisches Transferformat vorgeschlagen. Geeignet für diesen Zweck ist die *Extensible Markup Language* (XML). Ausgehend von der Annahme, dass lexikalische phonologische Informationen in Form von DATR (vgl. Carson-Berndsen/Walsh 2000), einer deklarativen Sprache zur Spezifikation nicht-monotoner Vererbungsnetze (vgl. Evans/Gazdar 1996), vorliegen, werden zunächst Aspekte der Transformation lexikalischen phonologischen Wissens vom zugrunde liegenden DATR- in ein XML-basiertes Format betrachtet. Daraufhin folgt eine Darstellung verschiedener Anwendungen, die multilinguale phonologische Repräsentationen in jeweils unterschiedlicher Qualität nutzen.

Maßgeblich für die Transformation lexikalischen Wissens ist es, einen Informationsverlust zwischen Eingabe- und Ausgabeformat zu vermeiden. Um eine Evaluation des hier vorgeschlagenen Ansatzes zu gewährleisten, wird im Folgenden die Natur phonologischer Beschreibungen im DATR-Format umrissen, um dann die Repräsentation in XML ableiten zu können.

DATR ist eine deklarative Sprache, die speziell zur Beschreibung natürlichsprachlicher Lexika entwickelt wurde. Zentral sind die Definitionen von Inferenz und verschiedenen Typen der Vererbung. DATR-Beschreibungen assoziieren Werte mit Knoten/Pfad-Paaren; dies geschieht entweder durch lokale Vererbung, durch globale Vererbung oder durch Defaults. In Anlehnung an Keller (1996) wird in den untenstehenden Abbildungen versucht, eine Logik für DATR Statements bereitzustellen. In DATR gibt es Formen expliziter und impliziter Vererbung. Während bei der expliziten Form zwischen lokaler und globaler Vererbung zu unterscheiden ist, wird die implizite Vererbung auch als Defaultvererbung bezeichnet. Lokale Vererbung basiert auf den vier folgenden Regeln:



Werte:	$\frac{}{\alpha \Rightarrow \alpha} \quad \text{Wert}$
Definitionen:	Wenn $K : \langle \alpha \rangle = \phi \in T$ , dann $\frac{\phi \Rightarrow \beta}{K : \langle \alpha \rangle \Rightarrow \beta} \quad \text{Def}$
Sequenzen:	$\frac{\phi \Rightarrow \alpha \vee \psi \Rightarrow \beta}{\phi \vee \psi \Rightarrow \alpha \beta} \quad \text{Seq}$
Evaluable Pfade:	$\frac{\phi \Rightarrow \alpha \quad K : \langle \alpha \rangle \Rightarrow \beta}{K : \langle \phi \rangle \Rightarrow \beta} \quad \text{Sub}$

Abb. 7: Semantik für lokale Vererbung in DATR.

Demnach wird etwa Abb. 8 wie folgt gedeutet: der Wert des Pfads  $P_1$  am Knoten  $K_1$  erbt seinen Wert von Pfad  $P_2$  am Knoten  $K_2$  und der Wert von Pfad  $P_2$  am Knoten  $K_2$  ist  $\alpha$ . Daraus schließt der Inferenzmechanismus, dass der Wert von Pfad  $P_1$  am Knoten  $K_1$  ebenfalls  $\alpha$  ist.

$$\frac{K_1 : P_1 = K_2 : P_2, K_2 : P_2 = \alpha}{K_1 : P_1 = \alpha}$$

Abb. 8: Beispiel: Lokale Vererbung in DATR.

In der DTD (*document type definition*) eines XML-Dokuments ist diese Art von Vererbung mittels des IDREF-Attributs generell darstellbar: ein Attribut vom Typ IDREF verweist auf ein Attribut vom Typ ID eines Elements im Dokument. Auf diese Weise ist es i. Ggs. zu anderen Attributtypen auch möglich, beliebig oft auf ein Element zu referieren. Das in Abb. 9 gegebene Beispiel konzentriert sich auf diesen Aspekt (unter praktischen Gesichtspunkten wäre es sicherlich vorzuziehen, nur ein Element  $P$  für alle Pfade anzunehmen; gleiches gilt für das Knotenelement):

in der DTD:	<code>&lt;!ATTLIST K1 wert ID #REQUIRED&gt;</code> <code>&lt;!ATTLIST K2 wert IDREF #REQUIRED&gt;</code>
im XML-Dokument:	<code>&lt;P1&gt;</code> <code>&lt;K1 wert="alpha"/&gt;</code> <code>&lt;/P1&gt;</code> <code>&lt;P2&gt;</code> <code>&lt;K2 wert="alpha"/&gt;</code> <code>&lt;/P2&gt;</code>

Abb. 9: Lokale Vererbung in XML (mittels ID und IDREF)

Formal lassen sich DATR-Theorien als Chomsky-2 Grammatiken charakterisieren (vgl. Cahill/Carson-Berndsen/Gazdar 2000); diesen formalgrammatischen Typ teilen sie mit der zugrunde liegenden Struktur von XML-Dokumenten, den Dokumenttypdefinitionen (DTDs). DTDs werden in einer formalen Syntax kodiert, die genau festlegt, welche Elemente und Entitäten im jeweiligen XML-Dokument zugelassen sind, sowie welchen Inhalt und welche Attribute diese Elemente enthalten dürfen.

Obwohl mit der Transformation von Abb. 8 in Abb. 9 genau genommen die rekursive Struktur der DATR-Semantik aufgegeben wird, und Abb. 8 stattdessen zwei alternative Repräsentationen erhält (vgl. Sailer/Richter 2001:165), sind in den DTDs hierarchische Strukturen festlegbar, die die Dominanzbeziehungen aus DATR-Theorien abbilden. Zusätzlich zu den Dominanzbeziehungen ist der oben demonstrierte Aspekt lokaler Vererbung in XML darstellbar.

Im Gegensatz zur lokalen Vererbung sind globale Mechanismen der Vererbung nicht in XML kodierbar. Dies trifft auch auf die oben bereits erwähnte Defaultvererbung zu, da die in DTDs üblichen *default attribute values* in Analogie zu lokaler Vererbung in DATR (s.o.) als lokale Defaults betrachtet werden können: Defaultwerte werden für jedes Attribut (d.h. lokal) definiert. In DATR jedoch ist der Begriff der Defaultvererbung umfassender, da er globale Defaults für ganze linguistische Kategorien vorsieht. Die in Abb. 10 folgenden Definitionen legen fest, wie ausgehend von Knoten/Pfad-Angaben bzw. von einzelnen Knoten oder Pfaden in einem bestimmten Kontext  $C$  in DATR auf globaler Ebene vererbt wird:

$$\begin{array}{l} \text{Knoten/Pfad-Referenz:} \quad \frac{(K, \alpha) \vdash K : \langle \alpha \rangle \Rightarrow \beta}{C \vdash "K : \langle \alpha \rangle" \Rightarrow \beta} \quad Ref_1 \\ \text{Pfad-Referenz:} \quad \frac{(K, \alpha) \vdash K : \langle \alpha \rangle \Rightarrow \beta}{(K, \alpha') \vdash "\langle \alpha \rangle" \Rightarrow \beta} \quad Ref_2 \\ \text{Knoten-Referenz:} \quad \frac{(K, \alpha) \vdash K : \langle \alpha \rangle \Rightarrow \beta}{(K', \alpha) \vdash "K" \Rightarrow \beta} \quad Ref_3 \end{array}$$

Abb. 10: Semantik für globale Vererbung in DATR.

Am Beispiel der Knoten-Referenz (vgl. Abb. 4,  $Ref_2$ ) kann nun gezeigt werden, dass eine Dokumenttypdefinition globale Vererbungsbeziehungen nicht erfassen kann. Während  $Ref_2$  für die DATR-Semantik festlegt, dass ein Knoten " $K$ " im aktualisierten Kontext  $(K, \alpha)$  direkt über den lokalen Deskriptor  $K : \langle \alpha \rangle$  evaluiert wird, ist dies in XML ausgeschlossen. In XML kann eine Evaluation für einen Knoten nur über lokale Pfade (vgl. Abb. 7, *Sub*) geschehen, wodurch keine vergleichbare Generalisierung ausgedrückt werden kann.

Derartige Einschränkungen werden in Goddard (2000) nicht problematisiert, sie sind jedoch gerade unter dem Gesichtspunkt des Informationsverlusts von Bedeutung.

Die *Extensible Markup Language* ist wie oben demonstriert ausschließlich in der Lage, eine Teilmenge der in DATR kodierten Informationen zu erfassen. Anders gesagt, können in XML partielle linguistische Beschreibungen repräsentiert werden (etwa in Form von *attribute-value matrices*, vgl. auch Sailer/Richter 2001:162 f.), die Repräsentation lexikalischer (Vererbungs-)Eigenschaften zwischen linguistischen Kategorien (d.h. nicht-lokal definierten Werten) übersteigt jedoch die Möglichkeiten XML-basierter Kodierungen. In dieser Hinsicht gilt auch für XML wie für SGML (*standard generalized markup language*), dass keine Mittel verfügbar sind, paradigmatische Relationen abzubilden (vgl. hierzu Gibbon 2000:23).<sup>1</sup> Die folgenden Anwendungen setzen die beschriebenen Einschränkungen voraus.

## 6 Phonologische Information in XML

In lexikalischen Silbentemplates wird die interne Struktur der phonologischen Konstituente *Silbe* festgelegt, in den Silbeneinträgen hingegen werden für tatsächliche und potenzielle Kombinationen sprachspezifische Einträge angelegt. Die folgenden Abbildungen zeigen ein Beispiel für ein Silbentemplate, einen Segment- und einen Silbeneintrag sowie darauf folgend die transformierte Kodierung einer vollständigen Repräsentation in XML.

```
Syllable:
  <> == Null

  <onset>   == "<phon onset first>"
             "<phon onset second>"
             "<phon onset third>"
  <nucleus> == "<phon nucleus first>"
             "<phon nucleus second>"
  <coda>    == "<phon coda first>"
             "<phon coda second>"
             "<phon coda third>"
             "<phon coda fourth>"
             "<phon coda fifth>"

  <structure> == I[ <onset> II <peak> II <coda> I].
```

Abb. 11: Silbentemplate in DATR.

<sup>1</sup> Allerdings weisen Sailer/Richter (2001) auch für *attribute-value matrices* darauf hin, dass zugunsten einer portablen XML-Kodierung von AVM-Beschreibungen die rekursive Struktur dieser Matrizen aufgebrochen wird.

Während das obige Silbentemplate die Struktur wohlgeformter Silben für das Deutsche vorgibt, ist der nun folgende Segmenteintrag nicht sprachspezifisch.

```
C_m: <> == C
      <manner> == [nasal]
      <place> == [bilabial]
      <phonation> == [voiced]
      <segmental> == m.
```

Abb. 12: Segmenteintrag in DATR.

```
S_knii:<> == Phonology
      <phon onset first> == "C_k:<>"
      <onset first duration> == 'null'
      <phon onset second> == "C_n:<>"
      <onset second duration> == 'null'
      <phon nucleus first> == "V_ii:<>"
      <nucleus first duration> == 'null'.
```

Abb. 13: Silbeneintrag in DATR (für Deutsch /kni:/ *Knie*).

Ausgehend von diesen lexikalischen Informationen wird die Ausgabe, d.h. die komplette phonologische Spezifikation, nun in XML transformiert. Damit wird sie zugänglich für weitere (computer-)linguistische Anwendungen, die allein mit dem generischen Transferformat arbeiten können, ohne auf das zugrunde liegende DATR-Format Bezug nehmen zu müssen. In der Kodierung finden sich alle notwendigen Informationen wieder, die in der DATR-Repräsentation vorgegeben sind. Vererbungsrelationen wie etwa, dass alle silbeninitialen Konsonanten von einem Onsetknoten dominiert werden (vgl. festgelegt in der DATR-Definition des Silbentemplates), erfahren gemäß der Vorarbeiten im vorangehenden Abschnitt eine abweichende Darstellung in der Mark-up-Kodierung. In der untenstehenden Abbildung werden derartige Relationen lokal, d.h. für jedes Onsetsegment durch ein umschließendes <onset>-Tag angelegt.

```
<syllable>knii
  <onset type="first"> ... <onset type="second"> ...
  <nucleus type="first">
    <vsegment
      phonation="voiced" manner="vowellike"
      place="front" height="high"
      round="false" length="tense" duration="null">ii
    </vsegment>
  </nucleus>
</syllable>
```

Abb. 14: Silbeneintrag in XML (für Deutsch /kni:/ *Knie*).

```

<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
- <multilingual_timemap>
  <startState>s1</startState>
  <finalState>s2</finalState>
  - <transition>
    <sourceState>s1</sourceState>
    <destinationState>s4</destinationState>
    <phonemeTape>m</phonemeTape>
    <graphemeTape>m</graphemeTape>
    <allophoneTape>m</allophoneTape>
    - <featureTape>
      <featureDescription>voiced</featureDescription>
      <featureDescription>nasal</featureDescription>
      <featureDescription>bilabial</featureDescription>
    </featureTape>
    - <featureConstraintTape>
      <threshold>4</threshold>
      - <overlapConstraint>
        <ranking>1</ranking>
        <featureDescription>voiced</featureDescription>
        <featureDescription>nasal</featureDescription>
      </overlapConstraint>
      - <overlapConstraint>
        <ranking>3</ranking>
        <featureDescription>voiced</featureDescription>
        <featureDescription>bilabial</featureDescription>
      </overlapConstraint>
      - <overlapConstraint>
        <ranking>2</ranking>
        <featureDescription>bilabial</featureDescription>
        <featureDescription>nasal</featureDescription>
      </overlapConstraint>
    </featureConstraintTape>
    <frequencyTape>22</frequencyTape>
    <durationTape>67</durationTape>
    <probabilityTape>0.7</probabilityTape>
  </transition>
</multilingual_timemap>

```

Abb. 15: XML-Kodierung der *Multilingual Time Map* für /m/.

Informationen über den phonotaktischen Automaten, der die Wohlgeformtheit einer Silbe wie der im vorangehenden Beispiel definiert, sind nun auch in der *multilingual*

*time map* enthalten. Für das Beispiel /m/ werden etwa Start- und Endzustände des endlichen Automaten, sowie die benötigten Überlappungsconstraints definiert.

Die Dokumenttypdefinitionen, die den logischen Strukturen der abgebildeten XML-Beispiele zugrunde liegen, werden automatisch aus den DATR-Informationen generiert. So kann etwa die Wertdeklaration in DATR (vgl. Abb. 16) exemplarisch für die übrigen Transformationen direkt in eine partielle Dokumenttypdefinition überführt werden.

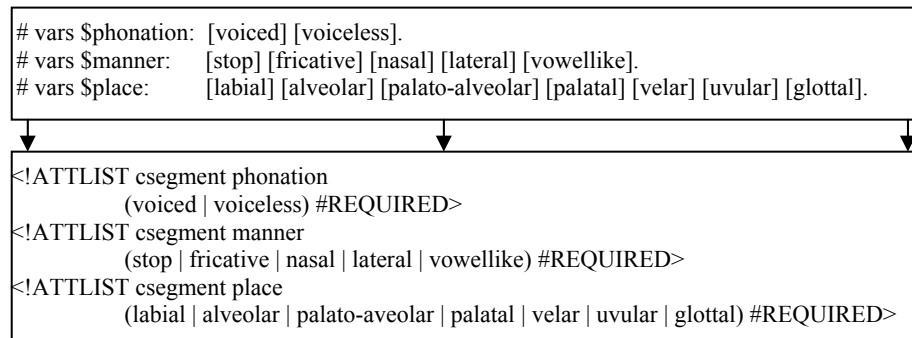


Abb. 16: Transformation von DATR-Statements in eine XML-basierte Attributliste.

Ein wichtiger Aspekt der *multilingual time map* ist die Transformation des generischen XML-Formats in idiosynkratische, benutzerabhängige Darstellungsformate. Mittels XPath, einer Syntax zur Beschreibung bestimmter Knoten oder Gruppen in einem XML-Dokument, können durch den Benutzer Repräsentationen unterschiedlicher Granularität angefordert werden. Mittels Transformationen, die in der *extensible stylesheet language* (XSL) definiert werden, ist zudem die Qualität für den Benutzer beeinflussbar. Die Umwandlung wird mithilfe mehrbändiger endlicher Automaten (*finite state transducer*) bewerkstelligt, die die in XML kodierten Daten in ein Multi-format überführen (darunter fallen sowohl Silbenstrukturen und multilineare merkmalsorientierte Listenrepräsentation, als auch phonetische Transkription in SAMPA). Die folgende Abbildung zeigt eine Auswahl möglicher Ausgaben:

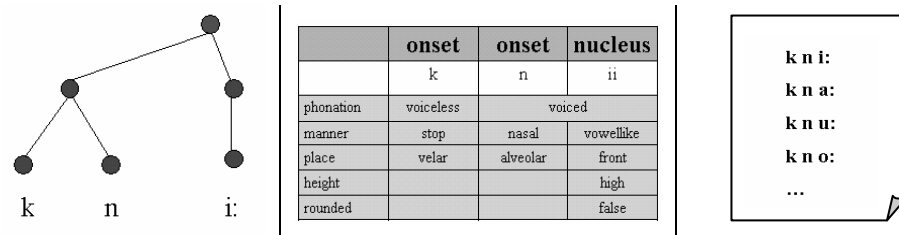


Abb. 17: Ausgabevarianten der phonologischen Informationen im Multifformat.

## 7 Ausblick

Auf der Grundlage der XML-Technologie wurde im vorangehenden Abschnitt die grundlegende Struktur multilingualer phonologischer Repräsentationen anhand des Konzepts der *multilingual time map* präsentiert.

Nach der Diskussion formaler Aspekte der Lexikontransformation konnte dargestellt werden, dass es nun einerseits mittels mehrbändiger endlicher Automaten möglich ist, phonologische Repräsentationen unterschiedlicher Qualität zu generieren. Zudem eröffnet das neuartige Format auch Möglichkeiten des hier nicht betrachteten Informationsaustauschs an der Syntax/Phonologie-Schnittstelle, strikt auf der Basis XML-konformer Kodierungen für AVM-Matrizen. Im HPSG-Bereich gibt es bereits eine Reihe von Datenbanken (Flickinger et al. 2000; Hinrichs et al. 2000), deren Ressourcen sich nun, da sie in ein einheitliches, portables Format transformierbar sind, verknüpfen lassen. Geeignet hierfür erscheinen mit Blick auf Sailer/Richter (2001) Attributwerte als Identifizierer vom Typ IDREF.

## Literatur

- Abu-Amer, T./Carson-Berndsen, J. (2003): A multilinear HMM based system for articulatory feature extraction. In Proceedings of ICASSP 2003, Hong Kong.
- Bohan, A./Creedon, E./Carson-Berndsen, J./Cummins, F. (2001): Application of a Computational Model of Phonology to Speech Synthesis, In: Proceedings of AICS2001, Maynooth.
- Cahill, L./Carson-Berndsen, J./Gazdar, G. (2000): Lexical Knowledge Representation. In: van Eynde, F./Gibbon, D. (Hrg.) *Lexicon Development for Speech and Language Processing*, Dordrecht: Kluwer Academic Publishers.
- Cahill, L./Gazdar, G. (1999): The PolyLex architecture: multilingual lexicons for related languages. *Traitement Automatique des Langues* 40 (2), 5-23.
- Carson-Berndsen, J. (1998): Time Map Phonology: Finite State Models and Event Logics in Speech Recognition. Dordrecht: Kluwer Academic Publishers.
- (2000): Finite State Models, Event Logics and Statistics in Speech Recognition. In: Sparck Jones, K./Gazdar, G. J. M./Needham, R. M. (Hrg.): *Computers, Language and Speech: formal theories and statistical data*. Philosophical Transactions of the Royal Society, Series A, Volume 358, issue no. 1769, 1255-1266.
  - (2002): Multilingual Time Maps; Portable Phonotactic Models for Speech Technology Applications. In: Proceedings of the LREC 2002 Workshop on Portability Issues in Human Language Technology Las Palmas, May 2002
- Carson-Berndsen, J./Walsh, M. (2000): Generic Techniques for Multilingual Speech Technology Applications. In: Proceedings of TALN 2000, Lausanne.
- Carson-Berndsen, J./Walsh, M. (in press): Phonetic Time Maps: Defining Constraints for Multilinear Speech Processing. In: van Dommelen/Barry (Hrg.): *Phonetic Knowledge in Speech Technology*. Dordrecht: Kluwer Academic Publishers.
- Carson-Berndsen, J./Gut, U./Kelly, R. (2003): Discovering Regularities in Non-native Speech. In Proceedings of Corpus Linguistics 2003, Lancaster.
- Evans, R./Gazdar, G. (1996): DATR: A language for lexical knowledge representation. In: *Computational Linguistics* 22, 2, 167-216.
- Flickinger, D./Copestake, A./Sag, I. A. (2000): HPSG Analysis of English. In: Wahlster, W. (Hrg.) *Verbmobil: Foundations of Speech-to-Speech Translation*. Berlin, Heidelberg, New York: Springer, S. 255-265.
- Gibbon, D. (2000): Computational Lexicography. In: van Eynde, F./Gibbon, D. (Hrg.) *Lexicon Development for Speech and Language Processing*. Dordrecht: Kluwer Academic Publishers, 1-42.
- Goddard, L. (2000): DATR in XML: DTD, Stylesheets, and Conversion Tools. Master's Thesis, University of Sussex.
- Goldsmith, J. (1990): *Autosegmental and Metrical Phonology*. Cambridge/Mass.: Basil Blackwell. Inc.
- Hinrichs, E. W./Bartels, J./Kawata, Y./Kordoni, V./Telljohann, H. (2000): The verbmobil treebanks. In: Proceedings of KONVENS 2000.
- Keller, Bill (1996): An Evaluation Semantics for DATR Theories. In: Proceedings of COLING 1996.



- Sailer, M./ Richter, F. (2001): Eine XML-Kodierung für AVM-Beschreibungen. In: Lobin, H. (Hrg.): Proceedings der GLDV-Frühjahrstagung 2001, 161-168.
- Tiberius, C. 2001. Multilingual Lexical Representation. PhD Thesis, ITRI, University of Brighton.
- Walsh, M./Wilson, S./Carson-Berndsen, J. (2002): XiSTS – XML in Speech Technology Systems. In: Proceedings of the COLING Workshop on XML & NLP, 67-74.