gefördert vom
Bundesministerium
für Forschung
und Technologie

Verbmobil
Verbundvorhaben

Schnittstellendefinition für den Worthypothesengraphen

E. Nöth, B. Plannerer

Friedrich-Alexander-Universität
Erlangen-Nürnberg
TU München

√m **Memo 2**
Dezember 1993

Dezember 1993

E. Nöth, B. Plannerer

TU München
Lehrstuhl für Datenverarbeitung
Franz-Joseph-Str. 38
D–80801 München

Lehrstuhl für Mustererkennung (Inf. 5)
Friedrich-Alexander-Universität Erlangen-Nürnberg
Martensstr. 3
D–91058 Erlangen

Tel.: (09131) 85 - 7888
e-mail: {noeth}@informatik.uni-erlangen.de

Gehört zum Antragsabschnitt: 3

Das diesem Bericht zugrundeliegende Forschungsvorhaben wurde mit Mitteln des Bundesministers für Forschung und Technologie unter dem Förderkennzeichen 01 IV 102 H/0 und 01 IV 102 C 6 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

1 Vorbemerkungen

In diesem Memo wird die Definition der Offline-Schnittstelle zwischen den Modulen "Akustik-Phonetik" und "Linguistik" zusammengestellt. Es handelt sich um eine Überarbeitung des ASL/Süd—Memos—7—92/TUM. Die wesentlichen Änderungen sind

- Die Festlegung des Formats für den Infostring.
- Definition der Schnittstelle als **zusammenhängender** Graph.
- Erweiterung der Schnittstelle dahingehend, daß eine Hypothese auch mehr als ein Wort überdecken kann.

Die Schnittstellendefinition wurde auf dem Verbmobil-Workshop *Vorverarbeitung und Spracherkennung* vom 21.11.1993-23.11.1993 in Blaubeuren von den dort anwesenden Gruppen verabschiedet.

Die Schnittstellendefinition legt das Format des Worthypothesengraphen fest, der von der Akustik-Phonetik an die Linguistik übergeben wird. Jede Kante des Graphen stellt eine bewertete Hypothese dar. In der Regel handelt es sich um Worthypothesen, es kann sich aber auch um größere Bereiche (z.B. haben_sie) und um nicht interpretierte Bereiche handeln (z.B. Telefonläuten). Da die Bewertungen der einzelnen Worthypothesen vermutlich je nach Akustik-Phonetik-Modul einen unterschiedlichen Wertebereich sowie eine unterschiedliche Interpretation (bedingte Emissionswahrscheinlichkeitsdichte, (Pseudo-)Rückschlußwahrscheinlichkeit auf einen Lexikoneintrag etc.) besitzen, soll hier KEIN einheitlicher Wertebereich vorgeschrieben werden. Vielmehr soll von den für die Akustik verantwortlichen Partnern jeweils eine entsprechende kurze Beschreibung des Bewertungsmasses geliefert werden. Fest vorgeschrieben sind also nur die in der Definition angegebenen Eigenschaften der Bewertungen.

2 Definition

2.1 Begriff "Offline-Schnittstelle"

Unter "Offline-Schnittstelle" soll in diesem Zusammenhang der Datenaustausch über Files oder verwandte Mechanismen (Pipe, Mailbox) verstanden werden. Geschwindigkeit und Datenmenge spielen hier eine untergeordnete Rolle gegenüber Konvertierbarkeit und Lesbarkeit. Daher kommen ausschließlich ASCII-Daten zum Einsatz.

2.2 Interaktion Linguistik-Akustik

Die Offline-Schnittstelle sieht vorläufig keine Interaktion zwischen den Modulen Linguistik und Akustik vor. Vielmehr beschreibt die Schnittstelle lediglich das Format, in dem die Akustik ihren Worthypothesengraph an die Linguistik weitergibt. Für eine echtzeitfähige Implementierung, die auch Interaktionen ermöglicht, wird eine neue Schnittstellendefinition erforderlich sein.

2.3 Schnittstellendefinition (18.2.1992, überarbeitet am 23.11.1993)

2.3.1 Bewertung der Worthypothesen

- Sei $K = W(1)...W(n-1)$ eine Teilkette mit der Bewertung $score(W(1)...W(n-1))$ und sei $W(n)$ eine hinzukommende Worthypothese mit der Bewertung $score(W(n))$, so ergibt sich für die Bewertung der Gesamtkette $G = W(1)...W(n)$:

  $$score(W(1)...W(n)) = score(W(1)...W(n-1)) + score(W(n)).$$

  und somit:

  $$score(G) = score(K) + score(W(n))$$

  Die Verknüpfung der Bewertungen erfolgt also additiv.

- Die Bewertungen verhalten sich ähnlich wie negativ logarithmierte Wahrscheinlichkeiten. Die in den Worthypothesengraphen eingetragenen Bewertungen sind also positive Zahlen, wobei die Sicherheit (Wahrscheinlichkeit) der Hypothesen zu größeren Zahlen hin kleiner wird. Der mögliche Zahlenbereich der Bewertung sowie eine eventuelle Normierung sind nicht festgelegt. Die für die Akustik-Phonetik verantwortlichen Partner sollen aber eine kurze Beschreibung ihres Bewertungsmasses liefern (Zahlenbereich, Interpretation, Normierung etc.).

2.3.2 Darstellung des Worthypothesengraphen

Der Worthypothesengraph ist **zusammenhängend**, d.h. jede Hypothese muß entlang eines Pfades vom Anfangsknoten aus erreichbar sein. Der Graph wird kantenweise geschrieben, wobei für jede Kante (und somit für jede Hypothese) eine eigene Zeile in folgendem ASCII-Format eingetragen und durch das < newline >-Symbol abgeschlossen wird:

Kante := <A E wort score ta te infostring>

dabei sind:

A:	logischer Anfangsknoten	(Kardinalzahl)
E:	logischer Endeknoten	(Kardinalzahl)
wort:	Hypothese	(String)
score:	Bewertung der Hypothese	(pos. Gleitkommazahl)
ta:	Anfangsframe	(Kardinalzahl)
te:	Endeframe	(Kardinalzahl)
infostring :	zusätzliche Informationen	(String)

Alle Einträge sind nach steigenden Anfangsknotennummern geordnet und innerhalb gleicher Anfangsknoten nach steigenden Endeknotennummern. Dabei dürfen Knotennummern nicht übersprungen werden. Der Graph beginnt mit der Anfangsknotennummer 1.

Die Knotennummern "A" und "E" müssen eine zeitliche Zuordnung ermögli-

chen, d.h., kleinere Knotennummern müssen sich auch auf zeitlich frühere Ereignisse beziehen. Die Schrittweite des durch die Knotennummern vorgegebenen Verarbeitungsrasters ist nicht festgelegt; es kann sich hierbei um ein Millisekunden-Raster, ein Frame-Raster oder ein Silben-Raster handeln.

Die Angaben "ta" und "te" sind Frame-Nummern. Diese werden vom Satzbeginn an fortlaufend gezählt. Der erste Frame eines Satzes hat die Nummer 1. Das entsprechende Zeitraster (in Millisekunden) ist von der jeweiligen Vorverarbeitung abhängig und nicht festgelegt. Die entsprechenden Umrechnungsformeln werden von den für die Akustik-Phonetik verantwortlichen Partnern geliefert.

Bei jeder Kante müssen zwingend die Informationen "A", "E", "wort", "score", "ta" und "te" angegeben werden.

Die Angabe "infostring" ist optional und kann zusätzliche Informationen (z.B. eine Zuordnung auf Laut-Ebene) und Informationen von anderen Wissensquellen (z.B. Prosodie) enthalten. Der "infostring" ist so aufgebaut, daß jede Wissensquelle darin etwas eintragen kann, und zwar vor das < *newline* >-Symbol. Jede Wissensquelle benutzt dabei folgendes Format:

(<Modulkennung> (<Info-Kennung> <Information>) ... (<Info-Kennung> <Information>))

Zur Zeit stehen folgende Kennungen fest:

- Prosodie
  (PR (G ...) (M ...) (A ...))
  - PR kennzeichnet, daß prosodische Information eingetragen wird.
  - G kennzeichnet, daß Information in Bezug auf die prosodische Markierung von Phrasengrenzen eingetragen wird.
  - M kennzeichnet, daß Information in Bezug auf die prosodische Markierung des Satzmodus eingetragen wird.
  - A kennzeichnet, daß Information in Bezug auf die prosodische Markierung von Phrasen- bzw. Satzakzent eingetragen wird.

  Nach dem Symbol G, M, bzw. A stehen N Bewertungen für die möglichen Klassen (z.B.

(G .1 .8 .1)

für die drei Klassen "keine Grenze", "schwache Grenze" und "starke Grenze" im Falle von G). Die Klassennamen werden nicht angegeben, um den Graphen kompakter zu halten. Anzahl und Reihenfolge der Klassen sowie Wertebereich der Bewertung wird, ebenso wie im Fall der Worthypothesen-Bewertung, in einer kurzen Beschreibung mitgeliefert.

- Akustik-Phonetik
  (AP (Z ...))
  - AP kennzeichnet, daß Zusatzinformation von der Worthypothesengenerierung eingetragen wird.
  - Z kennzeichnet, daß die zeitliche Zuordnung der Wortuntereinheiten eingetragen wird.
    Die Zuordnung hat die Anordnung "Wortuntereinheit" "Anfangsframe" für jede der verwendeten Untereinheiten. Der Endeframe ist durch den Anfangsframe der nachfolgenden Wortuntereinheit bzw. durch das Worthypothesenende bereits festgelegt und wird deshalb nicht angegeben. Die Untereinheiten (z.B. Phoneme oder Halbsilben) werden ebenfalls in einer kurzen Beschreibung festgelegt.

Die Worthypothesen "wort" entsprechen im wesentlichen den im Wortlexikon enthaltenen Einträgen, es erfolgt keine zusätzliche Umwandlung (Groß-/Kleinschreibung etc.) mehr. Es können auch aus den Einträgen im Wortlexikon größere Einheiten (Wortketten) gebildet und als Ganzes modelliert und hypothetisiert werden (z.B. um häufig auftretende starke Verschleifungen wie "ham' se" für "haben sie" besser erkennen zu können). In diesem Fall werden die einzelnen Wörter der Wortkette in ihrer kanonischen Form geschrieben und mit "_" verbunden, also "haben_sie" für obiges Beispiel.

Zusätzlich zu den Wortlexikon-Einträgen sind noch die Wörter "#PAUSE#", "#UW#" und "#NSE#" erlaubt. "#PAUSE#" kennzeichnet die Hypothese für eine Sprechpause. "#UW#" steht für "**u**nbekanntes **W**ort", also für einen von der Akustik-Phonetik nicht interpretierbaren sprachlichen Bereich. "#NSE#" steht für "**N**icht**S**prachliche **E**inheit", also für einen von der Akustik-Phonetik nicht interpretierbaren nichtsprachlichen Bereich wie z.B. das Klingeln eines Telefons. Auch für diese Hypothesen müssen alle nötigen Angaben (Bewertung etc.) gemacht werden.

Begrenzer zwischen den einzelnen Werten in einer Zeile sind Leerzeichen oder Tabulatorzeichen.

Beginn und Ende des Worthypothesengraphen werden durch die Schlüsselworte
"BEGIN_LATTICE" bzw. "END_LATTICE" gekennzeichnet, die jeweils in einer eigenen Zeile stehen müssen (siehe Beispiel).
Vor "BEGIN_LATTICE" bzw. nach "END_LATTICE" dürfen beliebig viele Kommentarzeilen in beliebigem Format eingetragen werden.

2.4 Beispiel:

```
Fiktiver Worthypothesengraph V1.1 erzeugt von E. Noeth
(dies ist ein Komentar-Bsp.)
BEGIN_LATTICE
1 2 #PAUSE# 11.05 1 110
2 3 ich 7.05 111 125 (AP (Z I 111 C 119)) (PR (G .95 .05) (M .9 .1 .0))
3 4 #NSE# 7.45 126 194
4 5 mu"s_am 8 195 208 (AP (Z m 195 u 198 s 201 a 205 m 206))(PR (G .3 .7)(M
.1 .8 .1))
4 6 kam 8.2 195 210 (AP (Z k 195 a 198 m 205))(PR (G .1 .8) (M .1 .8 .1))
6 7 #PAUSE# 11.7 211 891

(...)

END_LATTICE
```