

Wann geht der Sonderzug nach Pankow?

(Uhrzeitangaben und ihre prosodische Markierung in der Mensch-Mensch- und in der Mensch-Maschine-Kommunikation)

A. Batliner*, A. Kießling⁺, R. Kompe⁺, E. Nöth⁺, B. Raithel⁺

*: Institut für deutsche Philologie, Ludwig-Maximilians-Universität München

⁺: Lehrstuhl für Informatik 5 (Mustererkennung), Friedrich-Alexander-Universität Erlangen

1 Einleitung

Auskunftssysteme sind potentielle Anwendungen für die Mensch-Maschine-Kommunikation. Bei informationsabfragenden Dialogen im Szenario “telefonische Bahnauskunft” ist die Übermittlung von Abfahrts- und Ankunftszeiten der Züge von zentraler Bedeutung. Bei der schriftlichen Wiedergabe des folgenden Dialogs zwischen Bahnbeamten (B) und Kunden (K):

B: “Nürnberg ab 15 Uhr 23.” K: “15 Uhr 23./?”

etwa indizieren die Interpunktionszeichen (Punkt vs. Fragezeichen) den Unterschied zwischen Nicht-Frage und Frage. Bei gesprochener Sprache übernimmt in solchen Fällen die Prosodie (hauptsächlich die Intonation, d.h. der Tonverlauf) diese Rolle und steuert damit die Reaktion von B: bei einer Nicht-Frage ist keine spezielle Reaktion nötig, B kann im Dialog fortfahren; bei einer Frage muß B die Uhrzeit wiederholen. Auch wenn eine solche Behauptung unmittelbar einsichtig ist, kann nur die Untersuchung eines natürlichsprachlichen Korpus’ klären, ob eine solche, rein intonatorische Markierung bei elliptischen Uhrzeitangaben so häufig vorkommt, daß sie im System modelliert werden muß, und wie diese intonatorische Modellierung im Zusammenspiel mit der anderer Merkmale auszusehen hat.

2 Material

Das untersuchte Korpus besteht aus telefonischen Auskunftsdialogen, die bei der Reiseauskunft in Regensburg, Nürnberg, Erlangen und Wuppertal aufgenommen wurden. Die Aufnahmen in Erlangen und Wuppertal entstanden im Rahmen von [Kei90], die in Regensburg und Nürnberg sind dem FACID-Korpus entnommen, vgl. [Hit86]. Insgesamt standen 107 Dialoge zur Verfügung (Regensburg:14, Nürnberg:19, Erlangen:30, Wuppertal:44). Außer bei den Nürnberger Dialogen war der unwissende und damit auch unbeeinflusste Gesprächsteilnehmer immer der informationssuchende Kunde. Bei 92 dieser Dialoge geht es um Abfahrts- und Ankunftszeiten, bei den restlichen um andere Themen wie z.B. Preisauskünfte. In diesen 92 Dialogen finden sich 215 Äußerungen von K mit insgesamt 227 Wiederholungen der von ihm verstandenen Uhrzeit – bis auf drei Fälle bezieht sich diese Wiederholung auf die letzte von B genannte Uhrzeit. Es gibt drei mögliche illokutive Funktionen dieser Uhrzeitwiederholung, die nun in ihrer prototypischen Form charakterisiert werden sollen:

(1) Eine *Bestätigung* wird orthographisch durch einen Punkt wiedergegeben; die Intonation ist normalerweise fallend; Äquivalente zur Wiederholung der Uhrzeit sind Gliederungspartikeln wie “ja”, “mhm”, “aha”, etc. sowie die Wiederholung anderer wichtiger Information. Funktional entspricht die Bestätigung dem “Roger” im Funkverkehr und signalisiert damit Verstehen. Sie findet sich demnach oft nach dem Ende eines Redebeitrags (“turn”) von B, am Anfang der Redeübernahme (“turn taking”) durch K.

(2) Eine *Rückmeldung* kann orthographisch durch einen Bindestrich wiedergegeben werden; die finale Intonation ist normalerweise leicht steigend (“progreident”); Äquivalente sind ebenfalls Gliederungspartikeln bzw. die Wiederholung anderer wichtiger Information. Die Funktion kann am besten durch eine explizite Paraphrasierung in einem Dialog aus dem Korpus gezeigt werden:

B: 13 Uhr 18 ab Freiburg fahren

K: 13 Uhr 8-, ich schreib mir das einmal nur auf – 13 Uhr 18

Es wird also Zuhören und Verstehen signalisiert, zugleich kann aber das “Rederecht” von B kurzzeitig eingeschränkt werden, z.B. damit K die Uhrzeit aufschreiben kann. Eine Rückmeldung findet sich typischerweise als kurzer Einschub von K in einen Redebeitrag von B.

(3) Eine *Frage* wird orthographisch durch ein Fragezeichen wiedergegeben; der prototypische Intonationsverlauf ist steigend. Äquivalente sind Frageadverbien wie “*was*”, “*wie bitte*”, “*wann*”, etc. Funktional handelt es sich um eine Aufforderung zur Bestätigung der Richtigkeit oder zur Korrektur; solche Fragen finden sich oft, analog zu Rückmeldungen, als kurze Einschübe von **K** in einen Redebeitrag von **B**.

Die diskurssteuernde Funktion einer Bestätigung ist der einer Rückmeldung ähnlich, ihre intonatorische Form aber verschieden. Eine Bestätigung kann normalerweise gut von der Frage getrennt werden, die Unterscheidung von Rückmeldung vs. Frage (leicht steigend vs. steigend) kann Schwierigkeiten bereiten. Wie erwähnt, sind diese Charakterisierungen prototypisch und nicht erschöpfend; sie decken aber die meisten Uhrzeitwiederholungen ab.

Bei den 227 Wiederholungen von **K** reagiert **B** in rund 100 Fällen allein aufgrund der Intonation von **K** mit einer Bestätigung der Richtigkeit, einer Wiederholung oder einer Korrektur der Uhrzeit. In den anderen Fällen kann er sich nach anderen Indikatoren, wie z.B. Frageadverbien, richten. 64 mal tritt die Uhrzeit isoliert auf, d.h. **B** muß sich in diesen Fällen immer an der Intonation dieser elliptischen Uhrzeit orientieren. Im Schnitt treten also pro Dialog zwei Uhrzeitwiederholungen auf; einmal pro Dialog muß **B** aufgrund der Intonation von **K** reagieren, und in 2/3 dieser Fälle steht die von **K** geäußerte Uhrzeit allein. Damit bestätigt sich die Wichtigkeit der prosodischen Modellierung von Uhrzeitangaben im System.

3 Klassifizierung

In diesem Abschnitt soll dargestellt werden, welche Klassen unterschieden werden können, und auf welcher Ebene – in der Worterkennung oder mit Hilfe der Prosodie – die Reaktion von **B** gesteuert werden muß. Tab. 1 zeigt das Klassifikationsschema, Tab. 2 bringt zu jeder der Klassen ein Beispiel aus dem Korpus, soweit dies möglich ist.

	1	2	3	4	5	6	7	8	9a	9b	10	11	12
K wiederholt	<i>Richtig</i>			<i>Falsch</i> (Un-, + vollständig)			<i>Unvollständig</i> (aber richtig) <i>Stunden</i> <i>Minuten</i>						
Funktion	B	R	F	B	R	F	B	(R)	F		B	R	F
Intonation	↘	→	↗	↘	→	↗	↘	(→)	→ od. ↗		↘	→	↗
B reagiert	kR		b	k			kR	(kR)	e		kR		b
Erkennung der Notwendigkeit einer Reaktion des Systems durch:													
Worterkennung				XXX	XXX	XXX			XXX				XXX
Prosodie			XXX			XXX			XXX				XXX

Tab. 1: Klassifikationsschema

Die Numerierung der Klassen in Tab. 1 (erste Zeile) entspricht der Numerierung der Beispiele in Tab. 2. Die Wiederholung von **K** kann richtig sein, falsch oder unvollständig¹ (2. Zeile). Wenn sie falsch ist, muß **B** in jedem Fall korrigieren (vgl. ‘k’ in Zeile 5), egal, ob sie vollständig oder unvollständig ist. Bei der unvollständigen Wiederholung ist zu unterscheiden, ob nur die Stunden- oder nur die Minutenangabe wiederholt wurde. Für jede der drei Oberklassen von Zeile 2 gibt es drei unterschiedliche Funktionen, nämlich Bestätigung, Rückmeldung und Frage (‘B’, ‘R’, ‘F’ in Zeile 3) mit den dazu “passenden” Intonationsverläufen² (4. Zeile); zu anderen möglichen Verläufen vgl. weiter unten. Das System muß die Reaktion von **B** (5. Zeile) modellieren: Bei “richtig” reagiert **B** nur auf eine Frage mit einer Bestätigung (‘b’); sonst erfolgt keine Reaktion (‘kR’). Bei “falsch” muß **B** in jedem Fall korrigieren (‘k’). Bei “unvollständig” (aber richtig!) erfolgt im Prinzip bei Bestätigung und Rückmeldung keine Reaktion (‘kR’). Man beachte aber, daß es für die Klasse 8 im Korpus kein Beispiel gibt, vermutlich deswegen, weil in der Klasse 9 (Frage = Aufforderung zur Ergänzung) sowohl leicht steigender (9a in

¹Richtig, falsch, bzw. unvollständig bezieht sich auf die exakte Wiederholung der Uhrzeitangabe von **B** durch **K**. “21 Uhr” als Wiederholung von “21 Uhr” ist also vollständig, auch wenn es an sich eine unvollständige Uhrzeitangabe ist bzw. sein kann (quasi doppelt elliptisch für “21 Uhr null”).

² ↘: fallender, →: leicht steigender, ↗: steigender Intonationsverlauf.

Tab. 2) als auch steigender (9b in Tab. 2) Intonationsverlauf möglich sind, und bei leicht steigendem Verlauf die beiden Klassen 8 und 9a nicht unterschieden werden können.³ Bei Fragen (Klasse 9 und Klasse 12) muß ergänzt ('e') bzw. bestätigt ('b') werden.

Unter der Voraussetzung, daß die Worterkennung die korrekte Wortkette geliefert hat, kann die (durch 'XXX' in Tab. 1 angezeigte) Notwendigkeit einer Reaktion des Systems (Bestätigung, Ergänzung oder Korrektur) für die 3. Klasse nur von der Prosodie erkannt werden, in der 4. und 5. nur von der Worterkennung, und in der 6., 9. und 12. Klasse sowohl von Worterkennung wie auch von Prosodie. Bei den übrigen Klassen ist keine Reaktion notwendig.

1.	B: 10 Uhr 36 in Düsseldorf.		
	K:	10 Uhr 36.	Wunderbar.
2.	B: Dann fährt einer 16 Uhr 42		der ist in Schwerin 16 54.
	K:	16 Uhr 42 –	
3.	B: 12 Uhr 19 Hamburg Hauptbahnhof		äh, 12 Uhr 19, ja
	K:	12 Uhr 19?	
4.	B: Airport ist eine Station davor 8 Uhr 59		59
	K:	8 Uhr 55.	ja, und auch ...
5.	B: <i>Airport ist eine Station davor 8 Uhr 59</i>		59
	K:	8 Uhr 55 –	ja, und auch ...
6.	B: 21 Uhr.	21 Uhr.	Hält um 22 Uhr 3
	K:	22 Uhr?	21 Uhr. Aha! in Elberfeld?...
7.	B: ...11 Uhr 40 weiter.		
	K:	mhm. 11 Uhr.	
9a.	B: um 13 Uhr 19 ab Nürnberg, da können Sie mal durchfahren		19 ab Nürnberg
	K:	13 Uhr –	
9b.	B: ... 11 Uhr 50 in Neuss		50
	K:	äh, bitte? 11 Uhr?	11 Uhr 50 in Neuss, ja
10.	B:		16 Uhr 47
	K: Wann fährt der also in Frankfurt, der,	16 Uhr 37?	47.
11.	B: ... und dann 13 Uhr 22 ab Nürnberg		Laaber Hauptbahnhof 14 Uhr 33
	K:	22-	
12.	B: <i>21 Uhr 17 ab Hamburg.</i>		17, ja. Dann weiter...
	K:	17?	

Tab. 2: Beispiele (kursiv gedruckte Beispiele sind konstruiert, die relevanten Uhrzeitwiederholungen von K sind fett wiedergegeben)

4 Überlegungen zur Modellierung

Der einfachste Fall, auf den wir uns hier konzentrieren, sind isolierte, elliptische Uhrzeitangaben. Nicht alle möglichen Formen solcher Angaben treten häufig auf, einige finden sich im Korpus nicht. Tab. 3 zeigt die im Korpus möglichen Wortkombinationen und ihre Häufigkeit: Man sieht, daß die beiden vollständigen Kombinationen mit und ohne "Uhr" bei weitem die häufigsten sind; von den 19 *std*-*"Uhr"*-Kombinationen sind 5 im Sinne von Fußnote 1 vollständig. Die seltenen unvollständigen Angaben scheinen ein sprecherspezifisches Phänomen zu sein. In der Umgangssprache häufige Angaben wie *Behn nach drei* oder *viertel sechs* (regional) finden sich überhaupt nicht, da B sie nicht verwendet und K offensichtlich nicht die Angaben von B in Umgangssprache umkodiert. Auf der anderen Seite finden sich *"digitalisierte"* Angaben (vgl. *"15 null 9"* in Tab. 3) nur zweimal. Es bietet sich an, daß ein System grundsätzlich die häufigste Angabe mit "Uhr", die in ca. 80 % der Fälle zu beobachten ist, verwendet und damit auch K's Sprachgebrauch steuert.

Uhrzeitangaben stehen nicht immer allein: sie finden sich häufig zusammen mit Gliederungspartikeln (K: *7 33, ja*), Frageadverbien (K: *wie bitte? 7 33?*), Kommentaren (K: *7 33, das ist zu spät*) und

³Für die Klassen 5 und 12 sind die Beispiele konstruiert, da sich im Korpus keine fanden; u.E. gibt es dafür aber keine systematischen Gründe, wie es bei Klasse 8 der Fall zu sein scheint; deshalb ist in Tab. 2 für Klasse 8 kein Beispiel angegeben.

Wiederholungen von anderen Inhaltswörtern (K: 19 36 ab Elberfeld). Die Modellierung der Intonation (insbesondere die der terminalen Verläufe) und die Klassifizierung in Tab. 1 bleiben davon grundsätzlich unberührt, auch wenn berücksichtigt werden muß, ob solche Ergänzungen vor oder hinter der Uhrzeitangabe und damit präterminal oder terminal im Redebeitrag von K stehen. In der Worterkennung müssen diese Kombinationsmöglichkeiten natürlich mit erfaßt werden können. Bei der Intonation ist die Einteilung in *fallend*, *leicht-steigend*, *steigend* prototypisch und deckt die meisten Fälle ab, es gibt aber auch andere Intonationsverläufe: Wenn bei einer Bestätigung “9 Uhr 4!” der Hauptakzent auf “4” liegt, kann der Verlauf steigend sein; wenn bei einer gleichlautenden Frage der Hauptakzent auf “9” liegt, kann der Verlauf fallend sein, obwohl es sich um eine Frage handelt; vgl. zu diesen und weiteren Formen der Frageintonation und zur Möglichkeit ihrer Unterscheidung [Bat91]⁴. Diese “Ausnahmen” sollen später, ebenso wie die Auswirkungen der Silbenstruktur und der Akzentuierung auf den Tonverlauf, modelliert werden; man beachte, daß sie aber, im Gegensatz zu den prototypischen Verläufen, für den Menschen und damit auch für das System ambig sein können. Vom System sollte man in diesen Fällen sinnvollerweise nicht die Fähigkeit zur Disambiguierung verlangen – man muß es auch nicht, da eine Auflösung dieser Ambiguität nicht nötig ist: im Zweifelsfall können B bzw. das System einfach die korrekte Uhrzeit wiederholen bzw. mit “ja” bestätigen, ohne daß dadurch der Dialog gestört wird.

Aufbau	Anzahl	Beispiele
std “Uhr” min	162	7 Uhr 23
std min	36	8 15
std “Uhr”	19	12 Uhr
std	0	(14)
min	8	45
Ausnahmen	2	15 null 9 7 Uhr und 1

Tab. 3: Auftreten unterschiedlicher Uhrzeitstrukturen

5 Schlußbemerkungen

Die untersuchten Uhrzeitangaben sind zwar ein spezielles Phänomen des Szenarios “Bahnauskunft”, entsprechendes gibt es aber praktisch in allen Klärungsdialogen, in denen ein Teilnehmer wesentliche Information nachfragt bzw. bestätigt haben möchte und damit zugleich dialogsteuernd eingreift. In solchen Fällen wird sehr selten ein ‘ganzer, wohlgeformter Satz des Deutschen’ produziert; vielmehr werden entweder Inhaltswörter (Uhrzeit-/Ortsangaben bei der Reiseauskunft, Jahreszahlen/Seitenangaben bei Datenbankabfragen, Produktnamen/Verkaufszahlen/Preise bei Verkaufsgesprächen etc.) und/oder Gliederungspartikeln wie “*mhm*”, “*ja*”, “*aha*” etc. verwendet. Bei der Interpretation solcher elliptischer Einschübe spielt die Prosodie oft eine zentrale Rolle.

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministers für Forschung und Technologie unter den Förderkennzeichen 01IV102F4 und 01IV102H0 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literatur

- [Bat91] A. Batliner: *Ein einfaches Modell der Frageintonation und seine Folgen*. In E. Klein, F. P. Duteil, K. Wagner (Editoren): *Betriebslinguistik und Linguistikbetrieb*, S. 147–160, Niemeyer, Tübingen, 1991.
- [Hit86] L. Hitzenberger, R. Ulbrand, H. Kritzenberger, P. Wenzel: *FACID Fachsprachlicher Corpus informationsabfragender Dialoge*. FG Linguistische Informationswissenschaft Universität Regensburg, 1986.
- [Kei90] M. Keil: *Analyse von Partikeln für ein sprachverstehendes System – am Beispiel telefonischer Zugauskunftsdialoge*. Magisterarbeit, Philosophische Fakultät II, Universität Erlangen-Nürnberg, 1990.

⁴In dieser Untersuchung hat sich gezeigt, daß bei einem elizitiertem Korpus von ca. 2000 Äußerungen in rund 80% der Fälle die Höhe des finalen Grundfrequenzwertes (und damit implizit der terminale Intonationsverlauf) zur Unterscheidung von Frage und Nicht-Frage ausreicht. Dort wird auch auf die anderen Unterscheidungsmöglichkeiten sowie auf die Fälle, wo eine Neutralisierung vorliegt, d.h. nicht unterschieden werden kann, eingegangen.