In der Einheit über Signale, Nachrichten, ... haben wir auch über *Inschriften* gesprochen. Ein typisches Beispiel ist der Rosetta-Stein (Abb. 3.1), der für Jean-François Champollion *die* Hilfe war, um die Bedeutung ägyptischer Hieroglyphen zu entschlüsseln. Auf dem Stein findet man Texte in drei Schriften: in Hieroglyphen, in demotischer Schrift und auf Altgriechisch in griechischen Großbuchstaben.

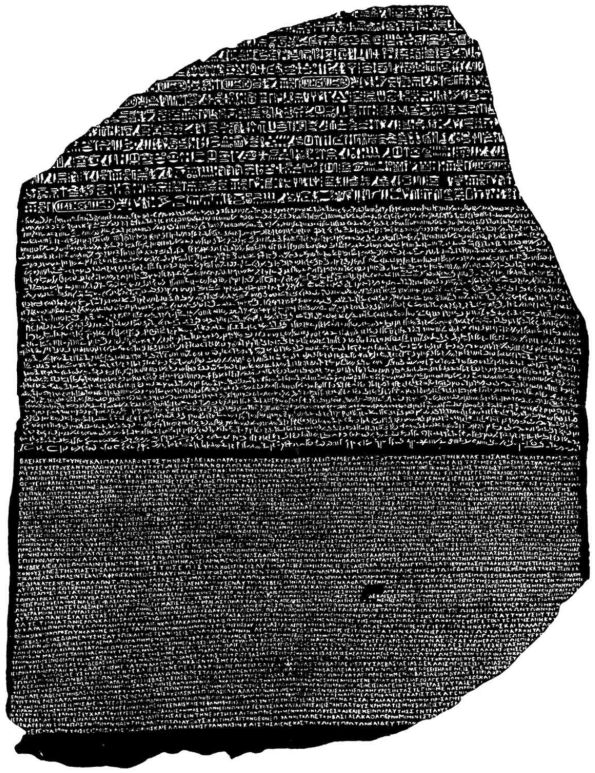

Abbildung 3.1: Der Rosetta-Stein, heute im Britischen Museum, London. Bildquelle: http://www.gutenberg.org/files/16352/16352-h/images/p1.jpg (19.10.08)

Wir sind gewohnt, lange Inschriften aus (Kopien der) immer wieder gleichen Zeichen zusammenzusetzen. Zum Beispiel in europäischen Schriften sind das die Buchstaben, aus denen Wörter aufgebaut sind. Im asiatischen Raum gibt es Schriften mit mehreren Tausend Zeichen, von denen viele jeweils für etwas stehen, was wir als Wort bezeichnen würden.

3.1 ALPHABETE

Alphabet

Unter einem *Alphabet* wollen wir eine endliche Menge sogenannter *Zeichen* oder *Symbole* verstehen, die nicht leer ist. Was dabei genau „Zeichen“ sind, wollen wir nicht weiter hinterfragen. Es

seien einfach die elementaren Bausteine, aus denen Inschriften zusammengesetzt sind. Hier sind einfache Beispiele:

- $A = \{\texttt{|}\}$
- $A = \{\texttt{a}, \texttt{b}, \texttt{c}\}$
- $A = \{\texttt{0}, \texttt{1}\}$
- Manchmal erfindet man auch Zeichen: $A = \{\texttt{1}, \texttt{0}, \texttt{ɪ}\}$
- $A = \{\texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{A}, \texttt{B}, \texttt{C}, \texttt{D}, \texttt{E}, \texttt{F}\}$

Gelegentlich nimmt man aber auch einen etwas abstrakteren Standpunkt ein und sieht zum Beispiel jeden der folgenden „Kästen“ als jeweils *ein* Zeichen eines gewissen Alphabetes an:

`int`
`adams`
`=`
`42`
`;`

3.1.1 *Beispiel ASCII*

Ein wichtiges Alphabet ist der sogenannte *ASCII*-Zeichensatz. Die Abkürzung steht für *American Standard Code for Information Interchange*. Diese Spezifikation umfasst insbesondere eine Liste von 94 „druckbaren“ und einem „unsichtbaren“ Zeichen, die man z. B. in Emails verwenden darf. Außerdem hat jedes Zeichen eine Nummer aus dem Bereich der natürlichen Zahlen zwischen 32 und 126. Die vollständige Liste findet man in Tabelle 1. Wie man dort sieht, fehlen in diesem Alphabet etliche Buchstaben aus nichtenglischen Alphabeten, wie zum Beispiel ä, ç, è, ğ, ñ, œ, ß, ů usw., von Kyrillisch, Japanisch und vielen anderen außereuropäischen Schriften ganz zu schweigen.

Auf ein Zeichen in Tabelle 1 sei ausdrücklich hingewiesen, nämlich das mit Nummer 32. Das ist das „Leerzeichen“. Man gibt es normalerweise auf einer Rechnertastatur ein, indem man die extrabreite Taste ohne Beschriftung drückt. Auf dem Bildschirm wird dafür in der Regel *nichts* dargestellt. Damit man es trotzdem sieht und um darauf aufmerksam zu machen, dass das ein Zeichen ist, ist es in der Tabelle als ␣ dargestellt.

3.1.2 *Beispiel Unicode*

Der Unicode Standard (siehe auch http://www.unicode.org) definiert mehrere Dinge. Das wichtigste und Ausgangspunkt für alles weitere ist eine umfassende Liste von Zeichen, die in der ein oder anderen der vielen heute gesprochenen Sprachen (z. B. in Europa, im mittleren Osten, oder in Asien) benutzt wird. Die Seite http://www.unicode.org/charts/ vermittelt einen ersten Eindruck von der existierenden Vielfalt.

Das ist mit anderen Worten ein Alphabet, und zwar ein großes: es umfasst rund 100 000 Zeichen.

		40	(	50	2	60	<	70	F
		41	)	51	3	61	=	71	G
32	␣	42	*	52	4	62	>	72	H
33	!	43	+	53	5	63	?	73	I
34	"	44	,	54	6	64	@	74	J
35	#	45	-	55	7	65	A	75	K
36	$	46	.	56	8	66	B	76	L
37	%	47	/	57	9	67	C	77	M
38	&	48	0	58	:	68	D	78	N
39	'	49	1	59	;	69	E	79	O
80	P	90	Z	100	d	110	n	120	x
81	Q	91	[	101	e	111	o	121	y
82	R	92	\	102	f	112	p	122	z
83	S	93	]	103	g	113	q	123	{
84	T	94	^	104	h	114	r	124	|
85	U	95	_	105	i	115	s	125	}
86	V	96	`	106	j	116	t	126	~
87	W	97	a	107	k	117	u		
88	X	98	b	108	l	118	v		
89	Y	99	c	109	m	119	w		

Tabelle 1: Die „druckbaren“ Zeichen des ASCII-Zeichensatzes (einschließlich Leerzeichen)

Der Unicode-Standard spezifiziert weitaus mehr als nur einen Zeichensatz. Für uns sind hier zunächst nur die beiden folgenden Aspekte wichtig[1]:

1. Es wird eine große (aber endliche) Menge A_U von Zeichen festgelegt, und
2. eine Nummerierung dieser Zeichen, jedenfalls in einem gewissen Sinne.

Punkt 1 ist klar. Hinter der Formulierung von Punkt 2 verbirgt sich genauer folgendes: Jedem Zeichen aus A_U ist eine nichtnegative ganze Zahl zugeordnet, der auch sogenannte *Code Point* des Zeichens. Die Liste der benutzten Code Points ist aber nicht „zusammenhängend“.

Aber jedenfalls liegt eine Beziehung zwischen Unicode-Zeichen und nichtnegativen ganzen Zahlen vor. Man spricht von einer Relation. (Wenn Ihnen die folgenden Zeilen schon etwas sagen: schön. Wenn nicht, gedulden Sie sich bis Abschnitt 3.2 wenige Zeilen weiter.)

Genauer liegt eine Abbildung $f : A_U \to \mathbb{N}_0$ vor. Sie ist

- eine Abbildung, weil jedem Zeichen nur *eine* Nummer zugewiesen wird,

[1] Hinzu kommen in Unicode noch viele andere Aspekte, wie etwa die Sortierreihenfolge von Buchstaben (im Schwedischen kommt zum Beispiel ö *nach* z, im Deutschen kommt ö *vor* z), Zuordnung von Groß- zu Kleinbuchstaben und umgekehrt (soweit existent), und vieles mehr.

- injektiv, weil verschiedenen Zeichen verschiedene Nummern zugewiesen werden,
- aber natürlich nicht surjektiv (weil A_U nur endlich viele Zeichen enthält).

Entsprechendes gilt natürlich auch für den ASCII-Zeichensatz.

3.2 Relationen und Abbildungen

Die Beziehung zwischen den Unicode-Zeichen in A_U und nichtnegativen ganzen Zahlen kann man durch die Angabe aller Paare (a, n), für die $a \in A_U$ ist und n der zu a gehörenden Code Point, vollständig beschreiben. Für die Menge U aller dieser Paare gilt also $U \subseteq A_U \times \mathbb{N}_0$.

Allgemein heißt $A \times B$ *kartesiches Produkt* der Mengen A und B. Es ist die Menge *aller* Paare (a, b) mit $a \in A$ und $b \in B$: *kartesiches Produkt*

$$A \times B = \{(a, b) \mid a \in A \wedge b \in B\}$$

Eine Teilmenge $R \subseteq A \times B$ heißt auch eine *Relation*. Manchmal sagt man noch genauer *binäre* Relation; und manchmal noch genauer „von A in B“. *Relation* *binäre Relation*

Die durch Unicode definierte Menge $U \subseteq A_U \times \mathbb{N}_0$ hat „besondere“ Eigenschaften, die nicht jede Relation hat. Diese (und ein paar andere) Eigenschaften wollen wir im folgenden kurz aufzählen und allgemein definieren:

1. Zum Beispiel gibt es für jedes Zeichen $a \in A_U$ (mindestens) ein $n \in \mathbb{N}_0$ mit $(a, n) \in U$.

   Allgemein nennt man eine Relation $R \subseteq A \times B$ *linkstotal*, wenn für jedes $a \in A$ ein $b \in B$ existiert mit $(a, b) \in R$. *linkstotal*

2. Für kein Zeichen $a \in A_U$ gibt es mehrere $n \in \mathbb{N}_0$ mit der Eigenschaft $(a, n) \in U$.

   Allgemein nennt man eine Relation $R \subseteq A \times B$ *rechtseindeutig*, wenn es für kein $a \in A$ zwei $b_1 \in B$ und $b_2 \in B$ mit $b_1 \neq b_2$ gibt, so dass sowohl $(a, b_1) \in R$ als auch $(a, b_2) \in R$ ist. *rechtseindeutig*

3. Relationen, die linkstotal und rechtseindeutig sind, kennen Sie auch unter anderen Namen: Man nennt sie *Abbildungen* oder auch *Funktionen* und man schreibt dann üblicherweise $R : A \to B$. *Abbildung* *Funktion*

   Gelegentlich ist es vorteilhaft, sich mit Relationen zu beschäftigen, von denen man nur weiß, dass sie rechtseindeutig sind. Sie nennt man manchmal *partielle Funktionen*. (Bei ihnen verzichtet man also auf die Linkstotalität.) *partielle Funktion*

4. Außerdem gibt es bei Unicode keine zwei verschiedene Zeichen a_1 und a_2, denen der gleiche Code Point zugeordnet ist.

   Eine Relation $R \subseteq A \times B$ heißt *linkseindeutig*, wenn für alle *linkseindeutig*

$(a_1, b_1) \in R$ und alle $(a_2, b_2) \in R$ gilt:

$$\text{wenn } a_1 \neq a_2, \text{ dann } b_1 \neq b_2 \,.$$

5. Eine Abbildung, die linkseindeutig ist, heißt *injektiv*. *injektiv*

6. Der Vollständigkeit halber definieren wir auch gleich noch, wann eine Relation $R \subseteq A \times B$ *rechtstotal* heißt: wenn für jedes $b \in B$ ein $a \in A$ existiert, für das $(a, b) \in R$ ist. *rechtstotal*

7. Eine Abbildung, die rechtstotal ist, heißt *surjektiv*. *surjektiv*

8. Eine Abbildung, die sowohl injektiv als auch surjektiv ist, heißt *bijektiv*. *bijektiv*

3.3 LOGISCHES

Im vorangegangenen Abschnitt stehen solche Dinge wie:

„Die Abbildung $U : A_U \to \mathbb{N}_0$ ist injektiv."

Das ist eine Aussage. Sie ist *wahr*.

„Die Abbildung $U : A_U \to \mathbb{N}_0$ ist surjektiv."

ist auch eine Aussage. Sie ist aber *falsch* und deswegen haben wir sie auch nicht getroffen.

Aussagen sind Sätze, die „objektiv" wahr oder falsch sind. Allerdings bedarf es dazu offensichtlich einer Interpretation der Zeichen, aus denen die zu Grunde liegende Nachricht zusammengesetzt ist.

Um einzusehen, dass es auch umgangssprachliche Sätze gibt, die nicht wahr oder falsch (sondern sinnlos) sind, mache man sich Gedanken zu Folgendem: „Ein Barbier ist ein Mann, der alle Männer rasiert, die sich nicht selbst rasieren." Man frage sich insbesondere, ob sich ein Barbier selbst rasiert ...

Und wir bauen zum Beispiel in dieser Vorlesung ganz massiv darauf, dass es keine Missverständnisse durch unterschiedliche Interpretationsmöglichkeiten gibt.

Das ist durchaus nicht selbstverständlich: Betrachten Sie das Zeichen $\mathbb{N}$. Das schreibt man üblicherweise für eine Menge von Zahlen. Aber ist bei dieser Menge die 0 dabei oder nicht? In der Literatur findet man beide Varianten (und zumindest für den Autor dieser Zeilen ist nicht erkennbar, dass eine deutlich häufiger vorkäme als die andere).

Häufig setzt man aus einfachen Aussagen, im folgenden kurz $\mathcal{A}$ und $\mathcal{B}$ genannt, kompliziertere auf eine der folgenden Arten zusammen:

NEGATION: Nicht $\mathcal{A}$.
Dafür schreiben wir auch kurz $\neg\mathcal{A}$.

LOGISCHES UND: $\mathcal{A}$ und $\mathcal{B}$.
Dafür schreiben wir auch kurz $\mathcal{A} \wedge \mathcal{B}$.

LOGISCHES ODER: $\mathcal{A}$ oder $\mathcal{B}$.
Dafür schreiben wir auch kurz $\mathcal{A} \vee \mathcal{B}$.

LOGISCHE IMPLIKATION: Wenn $\mathcal{A}$, dann $\mathcal{B}$.
Dafür schreiben wir auch kurz $\mathcal{A} \Rightarrow \mathcal{B}$.

Ob so eine zusammengesetzte Aussage wahr oder falsch ist, hängt dabei *nicht* vom konkreten Inhalt der Aussagen ab! Wesentlich ist nur, welche Wahrheitswerte die Aussagen $\mathcal{A}$ und $\mathcal{B}$ haben, wie in der folgenden Tabelle dargestellt. Deswegen beschränkt und beschäftigt man sich dann in der Aussagenlogik mit sogenannten *aussagenlogischen Formeln*, die nach obigen Regeln zusammengesetzt sind und bei denen statt elementarer Aussagen einfach *Aussagevariablen* stehen, die als Werte „wahr“ und „falsch“ sein können.

aussagenlogische Formeln

A	B	$\neg$A	A $\wedge$ B	A $\vee$ B	A $\Rightarrow$ B
falsch	falsch	wahr	falsch	falsch	wahr
falsch	wahr	wahr	falsch	wahr	wahr
wahr	falsch	falsch	falsch	wahr	falsch
wahr	wahr	falsch	wahr	wahr	wahr

Das meiste, was in obiger Tabelle zum Ausdruck gebracht wird, ist aus dem alltäglichen Leben vertraut. Nur auf wenige Punkte wollen wir explizit eingehen:

- Das „Oder“ ist „inklusiv“ (und nicht „exklusiv“): Wenn A und B beide wahr sind, dann auch A $\vee$ B.
- Man kann für komplizierte Aussagen anhand der obigen Tabellen „ausrechnen“, wenn sie wahr sind und wann falsch.

  Zum Beispiel ergibt einfaches Rechnen und scharfes Hinsehen, dass die Aussagen $\neg(A \vee B)$ und $(\neg A) \wedge (\neg B)$ immer gleichzeitig wahr bzw. falsch sind.

  Solche Aussagen nennt man *äquivalent*.

  äquivalente Aussagen
- Gleiches gilt für $\neg\,\neg A$ und A.
- Die Implikation $A \Rightarrow B$ ist auf jeden Fall wahr, wenn A falsch ist, unabhängig vom Wahrheitsgehalt von B, insbesondere auch dann, wenn B falsch ist. Zum Beispiel ist die Aussage

  Wenn $0 = 1$ ist, dann ist $42 = -42$.

  wahr.

  Man kann sich das so noch etwas klarer machen, dass man sich überlegt, was man sich denn unter dem „Gegenteil“ von $A \Rightarrow B$ vorstellen sollte. Das ist doch wohl $A \wedge \neg B$. Also ist $A \Rightarrow B$ äquivalent zu $\neg(A \wedge \neg B)$, und das ist nach obigem äquivalent zu $(\neg A) \vee (\neg\,\neg B)$ und das zu $(\neg A) \vee B$.
- Dabei haben wir jetzt so getan, als dürfe man selbstverständlich in einer Aussage einen Teil durch einen äquivalenten Teil ersetzen. Das darf man auch.

Alles was wir bisher in diesem Abschnitt betrachtet haben, gehört zu dem Bereich der *Aussagenlogik*. Wir werden sie vorläufig im beschriebenen Sinne naiv verwenden und in Zukunft zum Beispiel Definitionen wie die für Injektivität und Surjektivität von Funktionen entsprechend kompakter schreiben können.

Außerdem gibt es die sogenannte *Prädikatenlogik*. (Genauer gesagt interessiert uns Prädikatenlogik erster Stufe, ohne dass wir die Bedeutung dieser Bezeichnung jetzt genauer erklären wollen oder können.)

Aus der Prädikatenlogik werden wir — wieder zumindest vorläufig nur naiv — die sogenannten *Quantoren* verwenden: *Quantoren*

Allquantor $\forall$ *Existenzquantor* $\exists$

In der puren Form hat eine quantifizierte Aussage eine der Formen

$$\forall x\ A(x) \qquad \text{oder} \qquad \exists x\ A(x)$$

Dabei soll $A(x)$ eine Aussage sein, die von einer Variablen x abhängt (oder jedenfalls abhängen kann). A kann weitere Quantoren enthalten.

Die Formel $\forall x\ A(x)$ hat man zu lesen als: „Für alle x gilt: $A(x)$“. Und die Formel $\exists x\ A(x)$ hat man zu lesen als: „Es gibt ein x mit: $A(x)$“.

Zum Beispiel:

$$\forall x\ (x \in \mathbb{N}_0 \Rightarrow \exists y\ (y \in \mathbb{N}_0 \wedge y = x + 1))$$

Sehr häufig hat man wie in diesem Beispiel den Fall, dass eine Aussage nicht für alle x gilt, sondern nur für x aus einer gewissen Teilmenge M. Statt

$$\forall x\ (x \in M \Rightarrow B(x))$$

schreibt man oft kürzer

$$\forall x \in M :\ B(x)$$

wobei der Doppelpunkt nur die Lesbarkeit verbessern soll. Obiges Beispiel wird damit zu

$$\forall x \in \mathbb{N}_0 : \exists y \in \mathbb{N}_0 : y = x + 1$$

3.4 Zusammenfassung und Ausblick

In dieser Einheit wurde der Begriff *Alphabet*, eingeführt, die im weiteren Verlauf der Vorlesung noch eine große Rolle spielen werden. Die Begriffe *Wort* und *formale Sprache* werden gleich in der nächsten Einheit folgen.

Mehr über Schriften findet man zum Beispiel über die WWW-Seite http://www.omniglot.com/writing/ (1.10.08).

Als wichtige technische Hilfsmittel wurden die Begriffe *binäre Relation*, sowie *injektive*, *surjektive* und *bijektive Abbildungen* definiert, und es wurden informell einige Schreibweisen zur kompakten aber lesbaren Notation von Aussagen eingeführt.

Ein bisschen *Aussagenlogik* haben wir auch gemacht. Wir werden in einer späteren Einheit darauf zurück kommen, weil sich selbst hier schon schwierige algorithmische Problem verbergen.