Anhang B

Wie schreibt man Grammatiken?

Das Erstellen von Grammatiken ähnelt sehr dem Schreiben von Programmen. Die zugrunde liegende Aufgabe läßt sich sehr oft auf mehrere Arten lösen. Man sollte versuchen, durch Aufspalten des Problems in kleine, überschaubare Module die Wiederverwendung von bewährten Teillösungen zu fördern.

Das Aufstellen einer Grammatik geht in drei Schritten vor sich:

1. Identifizieren von *syntaktischen Kategorien*, d.h. von wichtigen Bestandteilen der zu beschreibenden Sprache. Diesen Kategorien werden Nonterminale zugeordnet.

2. Suchen von Beziehungen zwischen syntaktischen Kategorien. Diese Beziehungen werden dann mit Grammatikregeln beschrieben. Eine Grammatikregel zeigt, wie sich die syntaktische Kategorie auf der linken Seite aus den syntaktischen Kategorien der rechten Seite zusammensetzt. Verschiedene Regeln mit gleicher linker Seite entsprechen verschiedenen Spezialisierungen der syntaktischen Kategorie der linken Seite.

3. Aus der *abstrakten Syntax* (das Ergebnis der ersten beiden Schritte) wird nun die *konkrete Syntax* festgelegt. Das betrifft vor allem die Reihenfolge der Symbole auf der rechten Seite einer Produktionsregel und eventuell zusätzliche Terminalsymbole, um die Struktur zu betonen und die Lesbarkeit zu verbessern.

Als Beispiel wollen wir eine Grammatik aufstellen, die Zahlen in verschiedenen Zahlensystemen beschreiben soll.

1. Schritt: Als Kandidaten für syntaktische Kategorien fallen die Basis und die Ziffernfolge ins Auge. Eine Ziffernfolge besteht naturgemäß aus Ziffern bzw. Einzelbuchstaben. Wir führen als Bezeichnungen die Nonterminale `B`, `F` und `Z` ein. Den ganzen Ausdruck bezeichnen wir mit `A`. Der

Aufbau der Basis soll im Gegensatz zu den Ziffern der Ziffernfolge in der Grammatik nicht erklärt werden sondern als Einheit mit einem terminalen Symbol `num` betrachtet werden. Ihr Wertebereich liegt zwischen 2 und 36.

2. Schritt: Die Beziehungen zwischen den syntaktischen Kategorien sind leicht zu erstellen:

1. Ein Ausdruck `A` besteht aus einer Ziffernfolge `F` und der Basis `B`.
2. Eine Ziffernfolge `F` ist eine Aneinanderreihung von Ziffern `Z`.
3. Eine Ziffer ist eines der Zeichen '0' bis '9' oder einer der Buchstaben 'A' bis 'Z' (für Zahlensysteme bis zur Basis 36).
4. Eine Basis `B` ist eine Zahl `num`.

3. Schritt: Wir setzen die in den vorigen Schritten aufgestellte abstrakte Syntax aus Symbolen und Beziehungen nacheinander in konkrete Grammatikregeln um. Dabei wird die konkrete Syntax der Sprache festgelegt.

In unserem Beispiel muß die Ziffernfolge irgendwie von der Basis getrennt werden, damit wir diese beiden Gruppen von Zeichen auseinander halten können. Als Trennzeichen wählen wir das Zeichen '/'. Danach steht die Basis des Zahlensystems, z.B. A6E2/16, 10011/2, 755/8. Aus der ersten Beziehung ergibt sich die Regel

```
A  →  F '/' B
```

Die Ziffernfolge aus der zweiten Beziehung wird z.B. durch

```
F  →  F Z
F  →  Z
```

dargestellt.

Eine einfache Umsetzung der dritten Beziehung nimmt keine Rücksicht auf Klassen von Zeichen sondern listet ganz einfach alle Alternativen auf.

```
Z  →  '0' | '1' |...| '9' | 'A' | 'B' |...| 'Z'
```

Die letzte Beziehung ergibt sich zu

```
B  →  num
```

Die vollständige Grammatik hat damit die Form:

```
A  →  F '/' B
F  →  F Z
F  →  Z
Z  →  '0' | '1' |...| '9' | 'A' | 'B' |...| 'Z'
B  →  num
```

B.1 Qualitätskriterien für Grammatiken

Es gibt keine eindeutigen Kriterien für die Qualität von Grammatiken, wohl aber nützliche Stilregeln. Einige solcher Stilregeln sind:

- *Die Grammatik muß aufgabenmäßig abgefaßt sein.* Die syntaktischen Kategorien, die für die Verarbeitung von Bedeutung sind, müssen in der Grammatik festgelegt sein. Sind sie das nicht, dann läßt sich die Grammatik nur schwer attributieren. Die Beschreibung arithmetischer Ausdrücke hat z.B. als Forderung die unterschiedliche Priorität von Operatoren (Punktrechnung vor Strichrechnung). Sind diese Unterscheidungen schon in der Grammatik repräsentiert, dann vereinfacht sich die Attributierung ganz erheblich.

- *Die Grammatik muß analysegeeignet sein.* Dieser Punkt trifft besonders auf Grammatiken zu, die später mit Compilergeneratoren bearbeitet werden sollen. Für manche Syntaxanalyseverfahren sind z.B. Grammatiken mit linksrekursiven Symbolen nicht geeignet.

- *Grammatiken sollen eindeutig sein.* Die Eindeutigkeit erleichtert sowohl die Syntaxanalyse als auch die Zuordnung von Bedeutung durch die Attributierung.

B.2 Wie attributiert man Grammatiken?

Für die Attributierung geht man von der kontextfreien Grammatik aus und behandelt die semantischen Kategorien und deren Beziehungen untereinander. Beim Aufstellen einer Attributierung geht man wie folgt vor:

1. Identifizieren von *semantischen Kategorien*, d.h. von Eigenschaften bzw. Werten, die für die beabsichtigte Auswertung eine Rolle spielen. Zur Benennung dieser semantischen Kategorien werden Attribute eingeführt. Diese sind entweder synthetisierte oder ererbte Attribute, je nachdem, ob die betrachtete Information im Syntaxbaum von unten nach oben oder in umgekehrter Richtung fließen soll.

Dieser erste Schritt beeinflußt die entstehende Attributierung am nachhaltigsten. Man muß sich an dieser Stelle genau überlegen, welche Informationen man benötigt und in welcher Weise diese repräsentiert werden sollen.

2. *Zuordnung* von semantischen zu syntaktischen Kategorien, d.h. die Zuordnung von den Attributen zu den Nonterminalen. Es kann vorkommen, daß Nonterminale als „Zwischenlager“ für bestimmte Attribute eingesetzt werden müssen, die dort nicht benötigt werden. Nur so ist es möglich, Informationen weiter nach oben oder unten durchzureichen.

3. Bestimmen der *Attributauswertungsregeln:* Für jedes synthetisierte Attribut eines Nonterminals der linken Seite einer Regel und für jedes ererbte Attribut eines Nonterminals der rechten Seite einer Regel muß es eine Auswertungsregel geben.

Bei der Formulierung der Auswertungsregeln kann es notwendig sein, weitere Attribute einzuführen oder bestehende Attribute weiteren Nonterminalen zuzuordnen. Dies führt dann zu Iteration der Schritte 1 bis 3.

Betrachten wir noch einmal unser Beispiel. Wir wollen diese Grammatik attributieren, um den Wert einer Zahl zu berechnen.

1. Schritt: Als Ergebnis der Attributauswertung erwarten wir den Dezimalwert der Zahl in einem Attribut `val`. Bei der Betrachtung der Grammatik sehen wir, daß dieses Attribut Informationen in Richtung des Startsymbols zurückliefert, es wird also synthetisiert. Für die Basis brauchen wir auch ein Attribut `base`. Die Information wird jedoch bei der Berechnung des Wertes einer Ziffer gebraucht und muß daher als ererbtes Attribut zur entsprechenden Regel übergeben werden.

2. Schritt: Nun erfolgt die Zuordnung der semantischen zu den syntaktischen Kategorien. Alle Nonterminale erhalten das synthetisierte Attribut `val`. Die Ziffernfolge `F` erbt zusätzlich noch das Attribut `base`.

3. Schritt: Als letzten Schritt müssen wir alle konkreten Attributauswertungsregeln angeben. Dabei ist zu beachten, daß nur Attribute verwendet werden die im jeweiligen Kontext auch definiert sind.

Wir beginnen mit den Definitionen der Regeln von unten, d.h. bei den Blättern. Die Reihenfolge ist rein willkürlich gewählt. Sie folgt nur dem Fluß der synthetisierten Attribute, die in unserem Beispiel häufiger auftreten.

Den trivialen Abhängigkeiten der Regeln

```
F  →  Z
Z  →  '0' | '1' |...| '9' | 'A' | 'B' |...| 'Z'
B  →  num
```

entsprechen häufig „Transferregeln“ wie

```
F.val  :=  Z.val
Z.val  :=  0
Z.val  :=  1
           ⋮
Z.val  :=  35
B.val  :=  value( num.x )
```

die Attributwerte unverändert weiterreichen bzw. bei Terminalsymbolen die Werte erst zur Verfügung stellen[1]. Alle synthetisierten Attribute der linken Seiten haben damit einen Wert erhalten. Auf den rechten Seiten gibt es keine ererbten Attribute, denen ein Wert zugewiesen werden müßte.

Die eigentliche Berechnung des Zahlenwertes geschieht in der Regel

$$F_1 \rightarrow F_2\ Z$$

Der Wert für jede neue Ziffernfolge ergibt sich aus der alten Ziffernfolge multipliziert mit der Basis des Zahlensystems. Zu diesem Wert wird die neue Ziffer addiert. Die Basis steht im ererbten Attribut F_1`.base`. Damit und mit den restlichen synthetisierten Attributen der rechten Seite können wir die Attributierungsregel für das einzige synthetisierte Attribut der linken Seite formulieren. Fehlt nur noch die Regel für das einzige ererbte Attribut auf der rechten Seite. F_2`.base` bekommt den Wert von F_1`.base`.

Wir erhalten

```
F1.val   :=  F2.val * F1.base + Z.val
F2.base  :=  F1.base
```

In der ersten Produktion müssen wir jetzt nur noch den Wert der Basis zur Verfügung stellen und die errechnete Zahl an das Startsymbol übergeben. Es ergeben sich die Regeln

```
A.val   :=  F.val
F.base  :=  B.val
```

[1] Z.B. übersetzt die Funktion `value()` die Textrepräsentation (die im Attribut `x` abgelegt ist) der Zahl des Symbols in einen numerischen Wert.

Die vollständige attributierte Grammatik hat damit das folgende Aussehen:

Produktion			Regeln		
A	$\rightarrow$	F '/' B	A.val	:=	F.val
			F.base	:=	B.val
F_1	$\rightarrow$	F_2 Z	F_1.val	:=	F_2.val * F_1.base
					+ Z.val
			F_2.base	:=	F_1.base
F	$\rightarrow$	Z	F.val	:=	Z.val
Z	$\rightarrow$	'0'	Z.val	:=	0
Z	$\rightarrow$	'1'	Z.val	:=	1
	$\vdots$			$\vdots$	
Z	$\rightarrow$	'Z'	Z.val	:=	35
B	$\rightarrow$	num	B.val	:=	value(num.x)

Der Abhängigkeitsgraph in Abb. B.1 veranschaulicht den Informationsfluß bei der Ableitung einer dreistelligen Zahl (ererbte Attribute stehen links, synthetisierte rechts vom Ableitungsknoten).

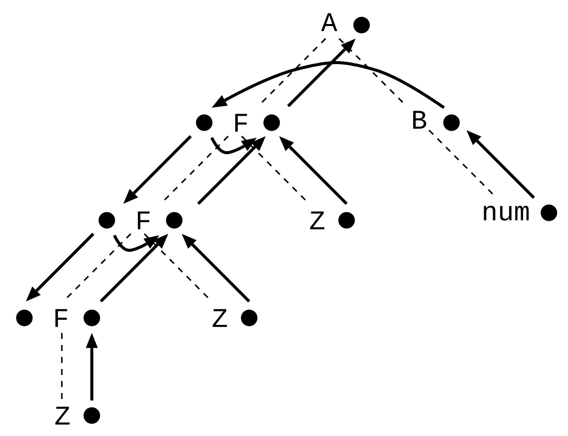

Abbildung B.1: Abhängigkeitsgraph

B.3 Qualitätskriterien für die Attributierung

Wir ergänzen die Stilregeln für Grammatiken um solche für Attributierungen:

- *Die Attributierung soll einfach und natürlich formuliert sein.* Dies empfiehlt sich wegen des Umfangs der Attributierung, der oft den der zugrundeliegenden Grammatik um ein Vielfaches übertrifft.

- *Die Attributierung muß vollständig sein*, d.h. in jedem Kontext muß für jedes zu berechnende Attribut eine Attributierungsregel zur Verfügung stehen.
- *Die Attributierung muß zyklenfrei sein.* Kein Attribut darf (direkt oder indirekt) von sich selbst abhängen.