

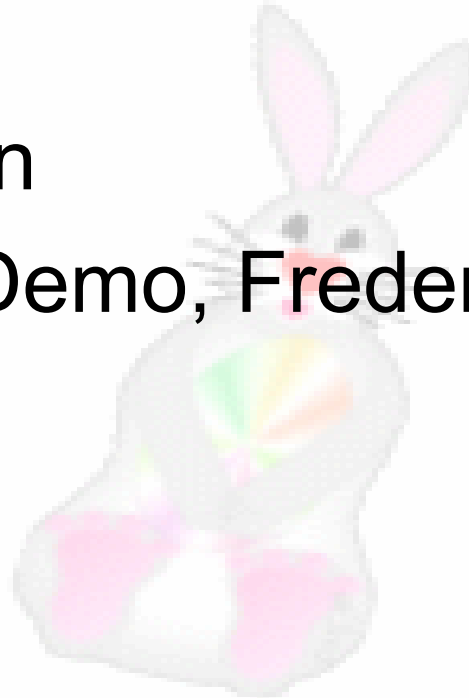
Practical Grammar Engineering Using HPSG 2.Tag

Frederik Fouvry, Petter Haugereid,
Valia Kordoni, Melanie Siegel



Inhalt

- Matrix
- Differenzlisten
- Debugging (Demo, Frederik)



Die LinGO Grammar Matrix

- Ein Nachteil von präzisen Grammatiken ist, dass der Aufbau aufwändig ist.
- Die Matrix ist ein „Starter Kit“, mit dem ein schneller Aufbau gewährleistet werden soll.
- Sie enthält Hypothesen über cross-linguale Eigenschaften von Sprache.

Einige Hypothesen der Matrix

- Wörter und Phrasen werden zu größeren Phrasen kombiniert.
- Die Semantik einer Phrase wird durch die Wörter und ihre Kombination bestimmt.
- Phrasenstrukturregeln können Semantik hinzufügen.
- Die meisten Phrasen haben eine identifizierbare Kopftochter.
- Köpfe bestimmen, welche Arten von Argumenten sie fordern und wie sie mit ihnen semantisch kombiniert werden.
- Modifikatoren bestimmen, welche Arten von Köpfen sie modifizieren und wie sie mit diesen semantisch kombiniert werden.
- Keine lexikalische oder syntaktische Regel löscht semantische Information.

Die Dateien in Matrix

matrix.tdl	enthält die Matrix-Typenhierarchie
irules.tdl	hier kommen Flexionsregeln rein
lrules.tdl	hier kommen lexikalische Regeln rein
lexicon.tdl	hier kommen die Lexikoneinträge rein
rules.tdl	hier kommen die Regeln rein
my_language.tdl	hier kommen die sprachspezifischen Typen rein
roots.tdl	hier werden die Startsymbole definiert
labels.tdl	hier bekommen die Knoten im Baum Namen
irregs.tab	für unregelmäßige Flexionen
README	Information über Matrix-Updates
Version.lsp	Information über die Matrix-Version

Die Dateien in Matrix

lkb/script	Definition der Dateien, die geladen werden
lkb/user-fns.lsp	Benutzereinstellungen zum Parsen
lkb/globals.lsp	Globale Parameter für die Grammatik
lkb/mrsglobals.lisp	Parameter für die Semantikkonstruktion
lkb/mrsmunge.tdl	Filter für semantisch leere Zeichen in der Generierung
pet.tdl und pet/	Hier stehen Einstellungen für die effiziente Verarbeitung mit dem PET System

Listen

- Manche Informationen in den Feature-Strukturen steht in Listen. Wir haben das z.B. schon bei ARGS gesehen:

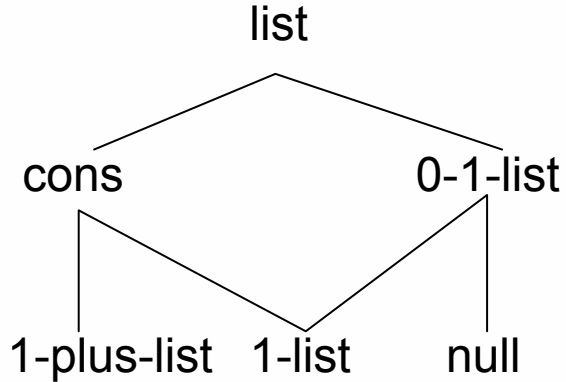
ARGS < phrase, phrase >

- TDL-Notation:

[FIRST phrase,
REST phrase]

Listen

- In matrix.tdl:



list := avm.

cons := list &
[FIRST *top*,
REST *top*].

0-1-list := list.

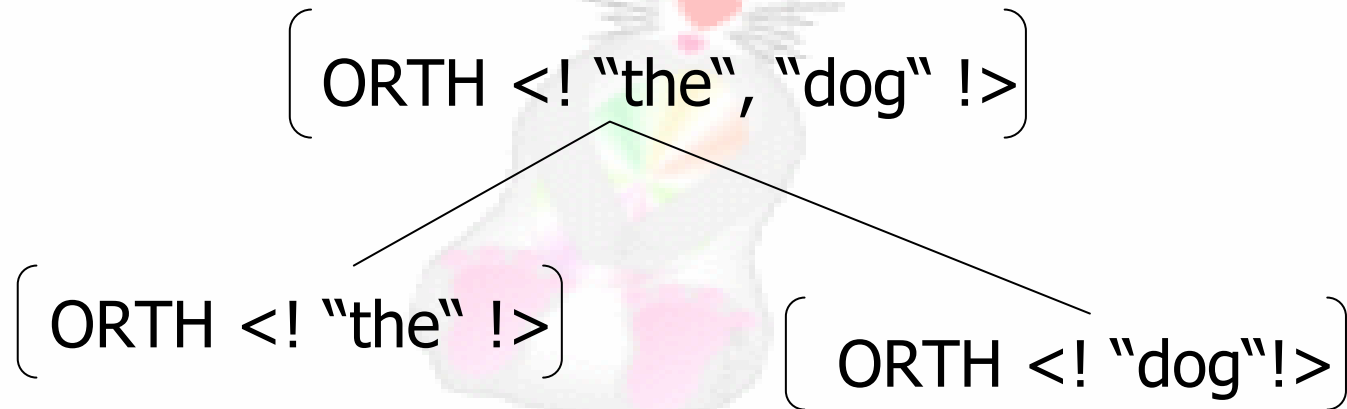
1-list := 0-1-list & cons &
[REST null].

null := 0-1-list.

1-plus-list := cons &
[REST cons].

Das Problem: Listen sollen zusammengefügt werden

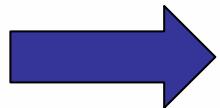
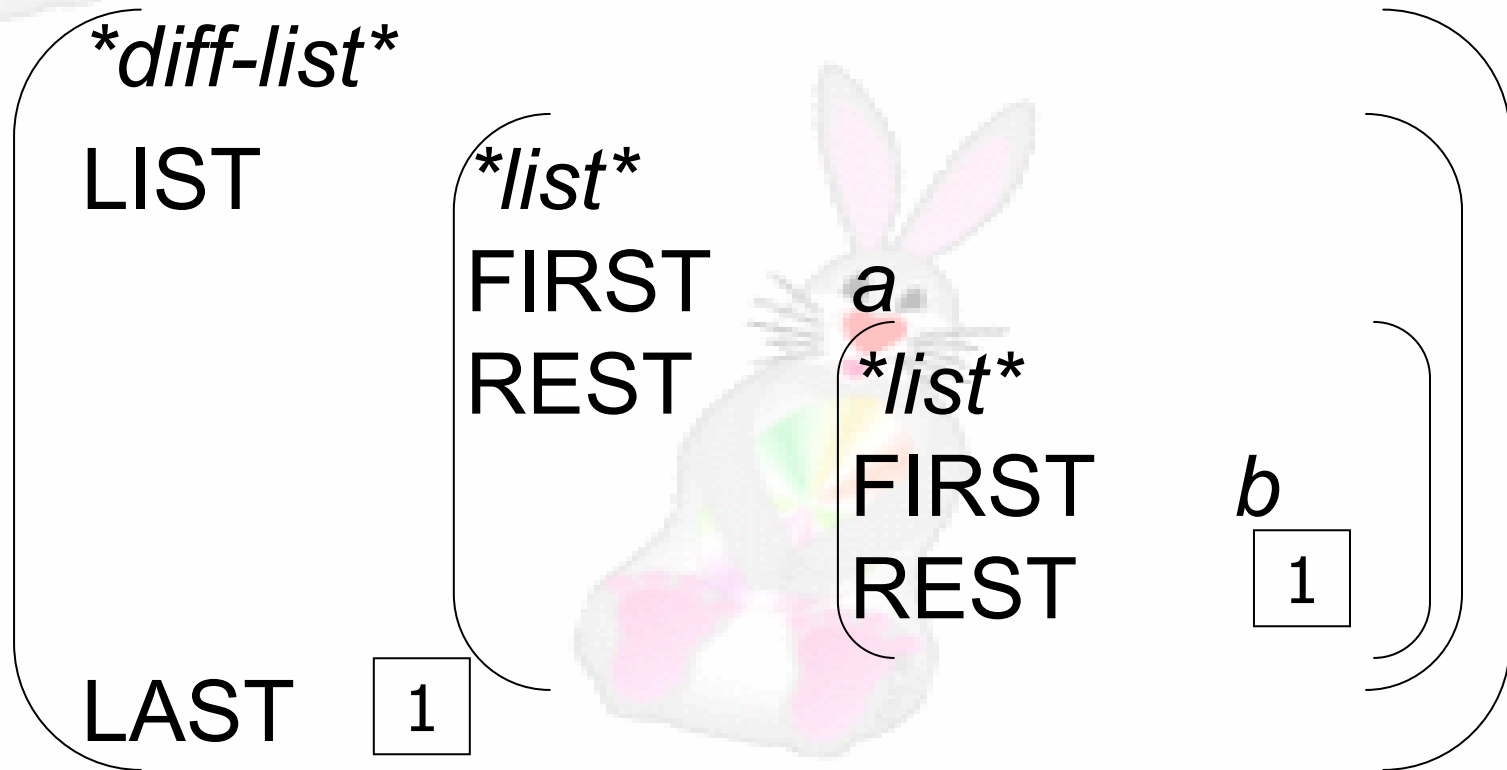
- Z.B.: Der Mutterknoten soll in ORTH eine Liste der abgearbeiteten Wörter enthalten.



Problematik der Listenzusammenfügung

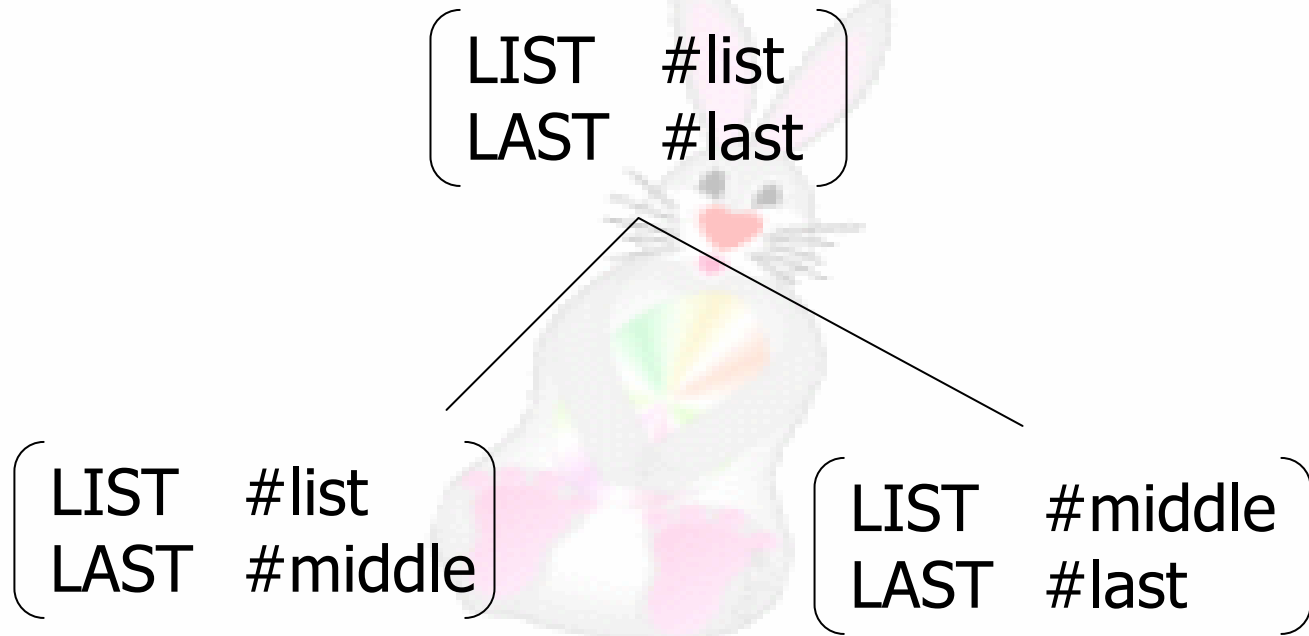
- Das Ergebnis soll möglichst wenig Struktur enthalten, am Besten einfach eine Liste sein.
- Dazu bräuchte man sowas wie *append*, aber dazu müsste man wissen, was das Ende der ersten Liste ist.
- Ein *append* von Listen mit bekannter Länge ließe sich definieren, aber ...

Die Lösung: Differenzlisten

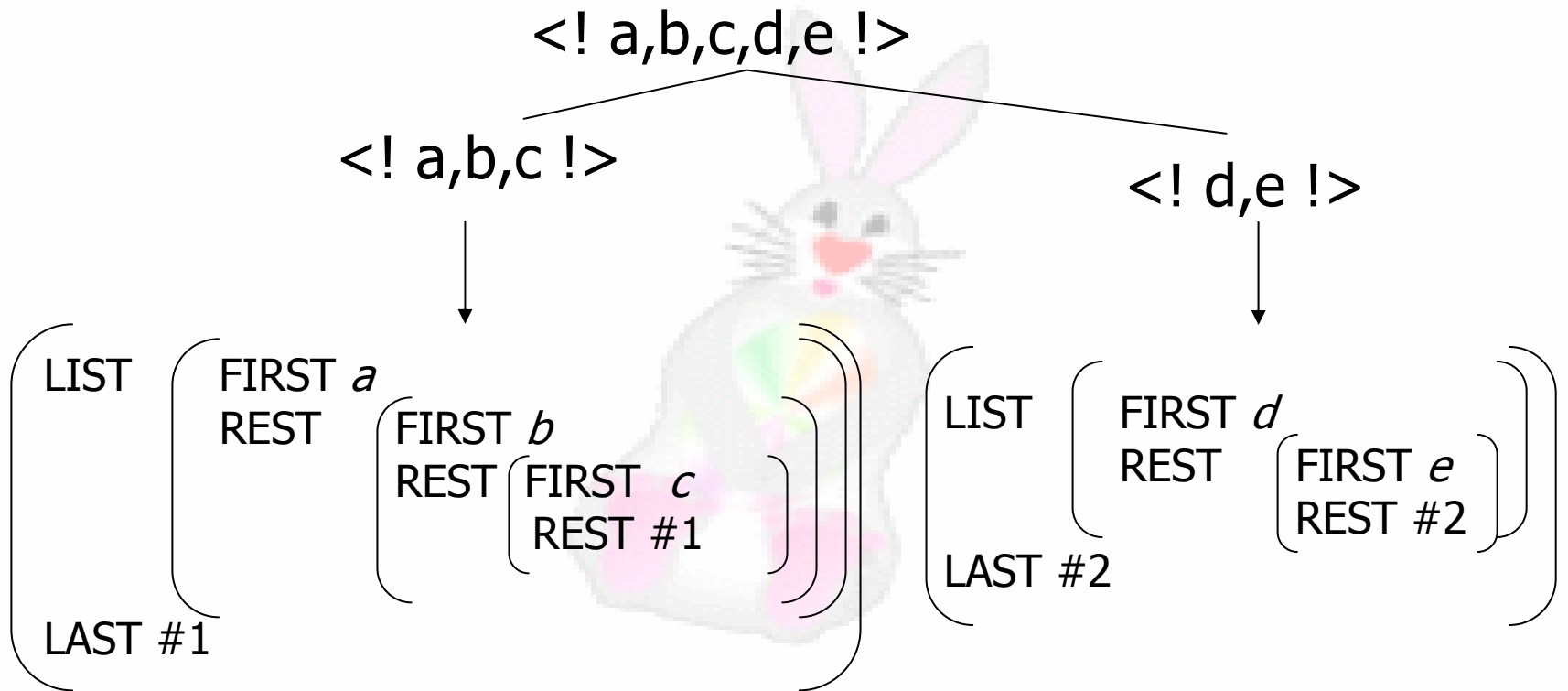


<! a , b !>

Einer Differenzliste etwas hinzufügen



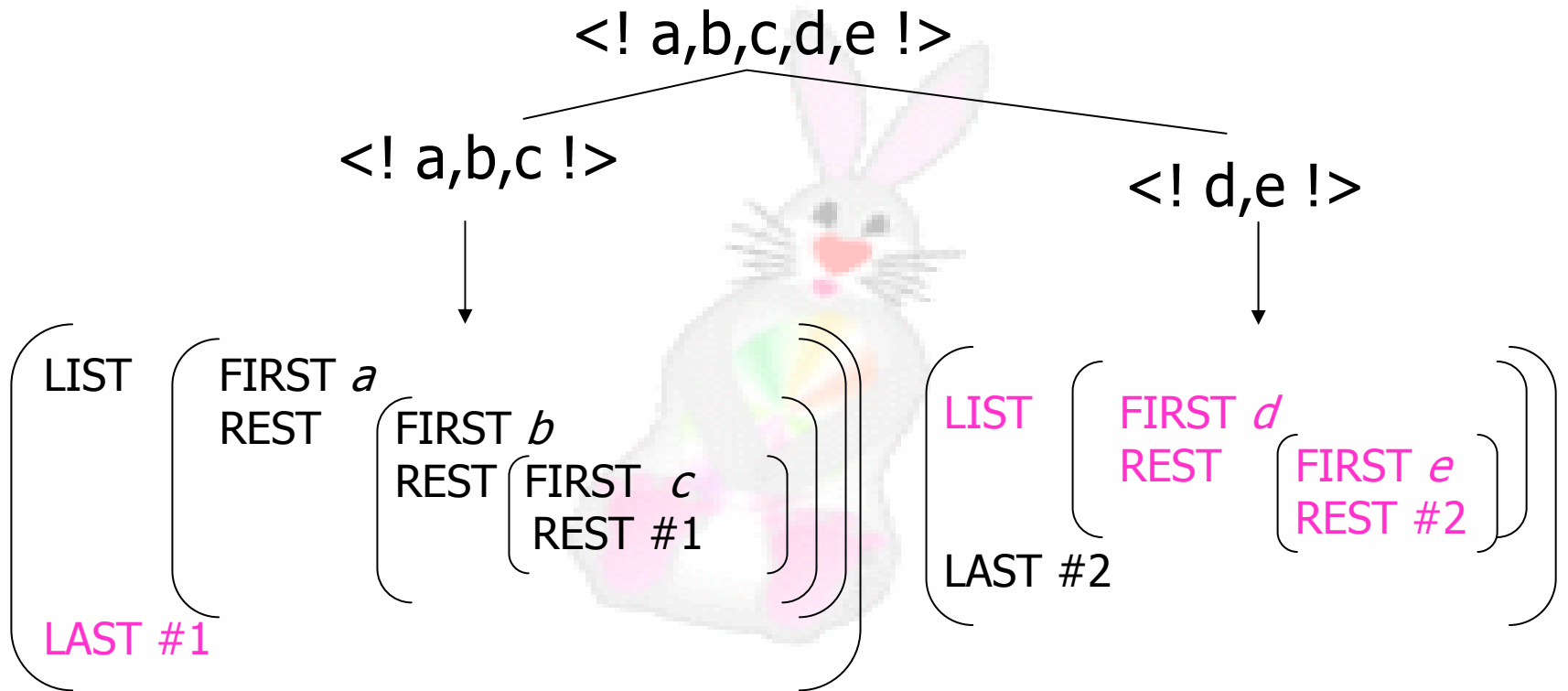
Zum Beispiel



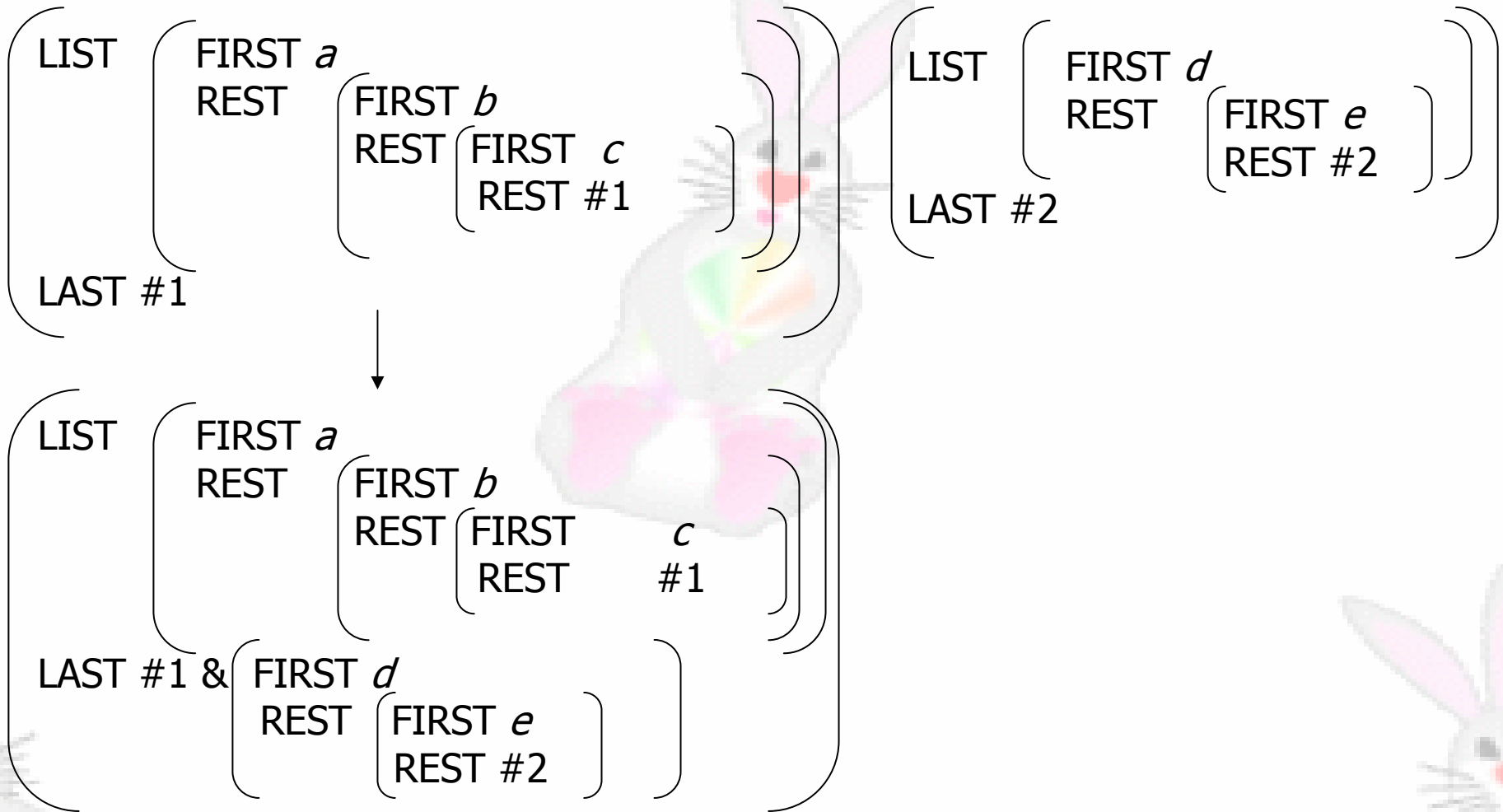
LAST der linken Tochter = LIST der rechten



LAST der linken Tochter = LIST der rechten



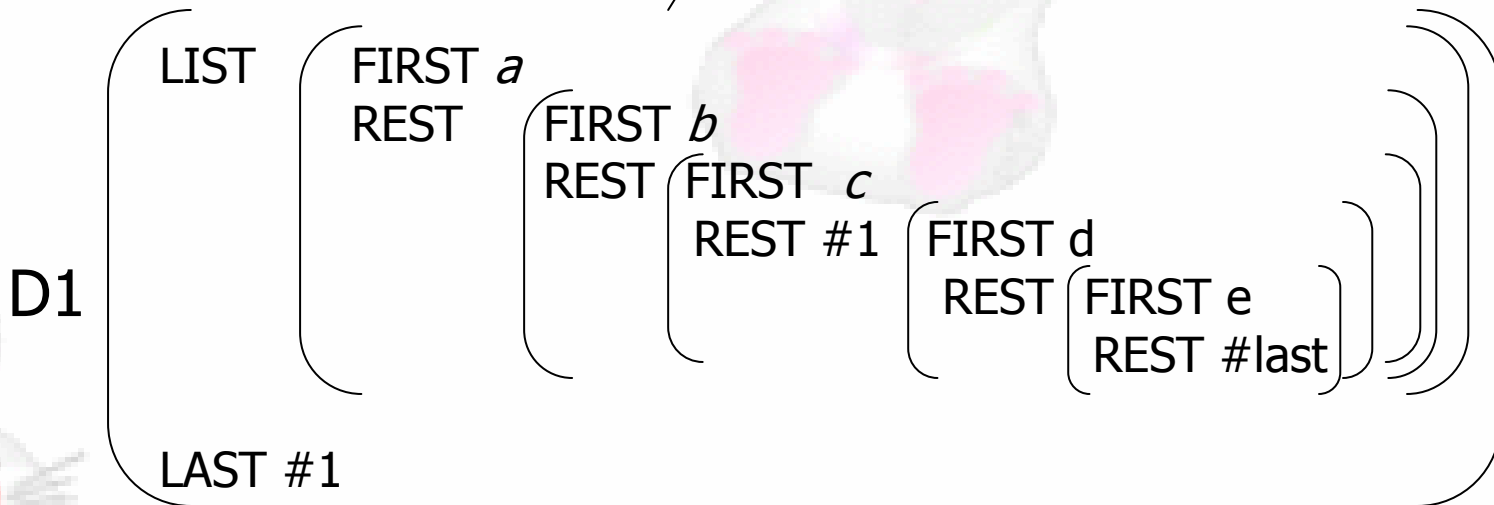
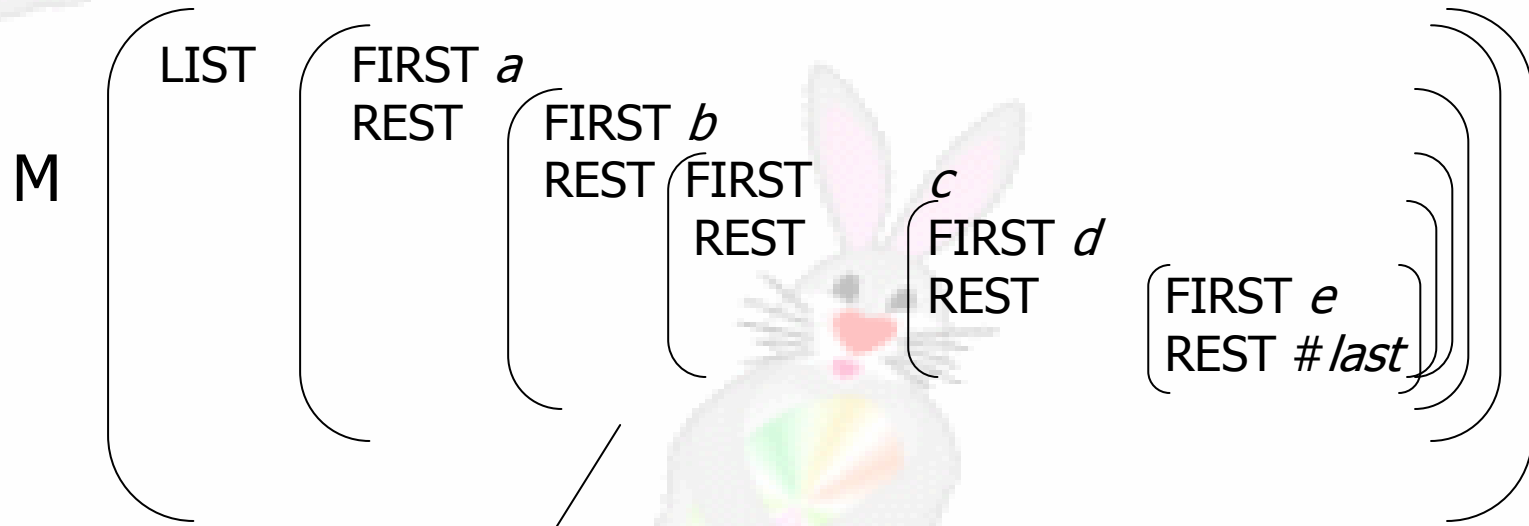
LAST der linken Tochter = LIST der rechten



LIST der Mutter = LIST der linken Tochter



LIST der Mutter = LIST der linken Tochter



Literatur und weitere Informationen

- On-Line User-Manual für LKB:
<http://www-csli.stanford.edu/~aac/lkb.html>
- Kursmaterialien:
<http://www.coli.uni-sb.de/~fouvry/grammatikentwicklung>
- Literatur:
 - Copestake, Ann: Implementing Typed Feature Structure Grammars. CSLI Publications, Stanford, CA (2002).
 - Copestake, Ann: The (New) LKB System. Manuscript. CSLI Stanford, Stanford, CA (2000). [<http://www-csli.stanford.edu/~aac/lkb.html>]
 - Sag, Ivan A., Thomas Wasow, and Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction*, 2nd edition. Stanford, CA: CSLI Publications.
 - Fouvry, Frederik and Oepen, Stephan: Introduction to Grammar Engineering Using HPSG. Scandinavian Summer School in Constraint-Based Grammar. Trondheim, Norge (2001). [<http://www.coli.uni-sb.de/trondheim01/>]
 - LinGO project, CSLI Stanford:
<http://lingo.stanford.edu>