

Klassifikation der nominalen Suffixe – ein computerlinguistisches Verfahren

Joanna Rabięga-Wiśniewska

Seminar für Computerlinguistik

Universität Warschau

jrw@mercury.ci.uw.edu.pl

Kolloquium für Examenskandidaten und Doktoranden

Prof. Dr. Tilman Berger

Slavisches Seminar, Tübingen, den 23. Juli 2003

Plan des Referats

1. Einführung

- a. Ziel der Arbeit
- b. Verfahren
- c. Empirisches Material

2. Beschreibungsmethode

- 1. Angaben in der linguistischen Literatur

3. Beschreibungsvoraussetzungen

- 1. Lexikon
- 2. Suffixsammlung

4. Lexikon

5. Klassifikation der Suffixe

- 1. polnische Suffixe
- 2. fremde Suffixe
- 3. semantische Information

6. Zusammenfassung

Ziel der Arbeit

- ein Derivationsmodell zu konstruieren
- praktische Anwendungen zu ermöglichen:
 - eine automatische morphologische (hier: derivative) Analyse und/oder Synthese polnischer Wörter;
 - Anlehnung an die Sammlung und Klassifizierung von formalen Komponenten des Motivierungsprozesses.

Ziel der Arbeit (2)

- Welche der Motivationsarten ist im Polnischen am häufigsten?
- Welches der Derivationsmittel ist am produktivsten?
- Welche Kombinationen von Derivationsmitteln sind möglich?
- Wie sieht das heutige polnische Wortbildungssystem aus?

Traditionelle Beschreibungsmethoden

- wichtige Informationen zum Derivat aus der Sicht der automatischen Sprachbearbeitung:
 - (a) besondere Flexionsinformationen,
 - (b) ausreichende thematische Alternationsbeschreibung,
 - (c) eindeutige Motivationswahl,
 - (d) formale Hinweise zur semantischen Kategorisierung.

Klassifikation der Suffixe

- Berücksichtigung folgender Punkte
 - eine eindeutige Flexionscharakteristik eines jeden Suffixes;
 - eine Bezeichnung der Flexionsklasse, die sie modifizieren;
 - Informationen darüber, ob sie Alternationen hervorrufen;
 - eine Beschreibung der Varianten, aber nur der häufig in Texten anzutreffenden;
 - und dann Informationen über ihre Fähigkeit, sich mit anderen Suffixen (und wenn ja, mit welchen) zu verbinden.

Die Beschreibungsmethode

Meine Hypothese lautet, dass die flexivischen Formen des Wortes eine Basis für die derivativen Formen bilden.

Vorteil:

- Alternationen können als innerer flexivischer Formenwechsel betrachtet werden
- viele detaillierte, aber unnötige Regeln können eingespart werden

Beispiel der Analyse

1.

ogrzew-acz \leftarrow ogrzew(ać)_{imperf} statt \leftarrow ogrz(ać)_{perf}
(ein Wärmer \leftarrow wärmen)

zszyw-acz \leftarrow zszyw(ać)_{imperf} statt \leftarrow zszy(ć)_{perf}
(ein Heftapparat \leftarrow zusammennähen)

2.

porównywal-ny \leftarrow porównywał_{3 sing prät} statt \leftarrow
porównywać_{inf} (Szymanek 1985)
(vergleichbar \leftarrow vergleichen)

widzial-ny \leftarrow widział_{3 sing prät} statt \leftarrow widzieć_{inf}
(sichtbar \leftarrow sehen)

Beispiel der Analyse (2)

3.

grzał-ka ← grzał_{3 sing prät} statt ← grzać_{inf}
(ein Tauchsieger ← wärmen)

pral-ka ← prał_{3 sing prät} statt ← prać_{inf}
(eine Waschmaschine ← waschen)

4.

grzejnik ← grzej(e)_{3 sing präs} statt ← grzać_{inf} (Grzegorzyczkowa
1979)

(eine Heizung ← heizen)

nadajnik ← nadaj(e)_{3 sing präs} statt ← nadawać_{inf}
(ein Sender ← senden)

Beschreibungsvoraussetzungen

Zwei Hauptvoraussetzungen:

- ein erforderlicher Zugang zu jeder Wortform der hypothetischen Basis – das Lexikon
- eine Sammlung der Suffixe (der Präfixe) – die Datenbank

Das Lexikon

- Grundlage – das grammatische Lexikon des Flexionsanalysators AMOR
- Beispiel: kocioł (*ein Kessel*)

mIva

9: N3=cioł G3=#a D3=#owi A3=cioł
B3=#em L3=#e V3=#e n3=#y g3=#ów
d3=#om a3=#y b3=#ami l3=#ach v3=#y
ko:9

Das Lexikon (2)

- Two Level Morphology Modell
- Jede lexikalische Einheit auf zwei Ebenen:
 - eine graphische Realisierung
 - ein phonetischer oder konventioneller Ausdruck

Die Unterschiede zwischen den Ebenen bestimmen die Regeln der kontextuellen Umbildungen.

Das Lexikon (3)

bosman r:1 bosmaN

{N:n(N,G,D,A,B,n,g,d,a,b,l,v),ni(L,V)} k(0, a, owi,
a, em, e, e, i, ów, om, ów, ami, ach, i)

(ein Bootsmann)

robot r:2 roboT

{T:t(N,G,D,A,B,n,g,d,a,b,l,v),ci(L,V)} k(0, a, owi,
a, em, e, e, y, ów, om, y, ami, ach, y)

(ein Roboter)

pomost r:3 pomoST

{S:s(N,G,D,A,B,n,g,d,a,b,l,v),ś(L,V);T:t(N,G,D,A,
B,n,g,d,a,b,l,v),ci(L,V)} k(0, u, owi, 0, em, e, e,
y, ów, om, y, ami, ach, y)

(ein Steg)

Das Lexikon (4)

studnia r:z stuD0N

{D:d(N,G,D,A,B,L,V,n,g,d,a,b,l,v),dzi(g);0:0(N,G,D,A,B,L,V,n,g,d,a,b,l,v),e(g);N:ni(N,A,B,V,n,g,d,a,b,l,v),n(G,D,L,g)} k(a, i, i, e, a, i, o, e, i.0, om, e, ami, ach, e)

(ein Brunnen)

przedzielony przedzielOn

{O:o(Npm,Npn,Npz,Gpmn,Gpz,Dpz,Apa,Apd,Apn,Apz,Bpmn,Lpz,npb,gp,dp,apo,apb,bp,lp),e(npoy.e.a, ego.ej, emu.ej, ego.y.e.a, ym.a, ym.ej, ^, i.e, ych, ym, ych.e, ymi, ych, ^)}

(geteilt)

Das Lexikon (5)

- Vorteile:
 1. Möglichkeit, die Alternationen zu gruppieren;
 2. Möglichkeit, die Abweichungen der Deklination zu untersuchen;
 3. Möglichkeit, jede flexivische Form des Lexems zu bilden.

Die Klassifikation der Suffixe

● Die polnischen Suffixe

- Grzegorzczkowa und Puzynina (1979)
- 109 Suffixe mit 253 Varianten

● Bedingungen zur Exzerption

- die Quantität von den mit einem Suffix abgeleiteten Derivaten lag zwischen 8000 und 80;
- es wurde angenommen, dass eine Motivation einen Typ der Derivate charakterisiert;
- bestimmte Suffixe können miteinander kombiniert werden;
- die Beschreibung der Suffixe kann die flexivische Charakteristik der Basis einschließen.

Die Klassifikation der Suffixe (2)

- Ergebnis: 29 Suffixe mit 20 Varianten

SUFFIX	ANZAHL
-ka	8000
-ość	3500
-ik	2200
-ek	2000

Die Klassifikation der Suffixe (3)

● Die fremden Suffixe

- Waszakowa (1994)
- 4300 Nomina
- 65 Suffixe mit 86 Varianten
- 2 Suffixe mit der Quantität über 500
- 2 Suffixe, die über 300 Wörter bilden
- 10 Suffixe, die 50-120 Wörter bilden

Die Klassifikation der Suffixe (4)

- Ergebnis: 14 Suffixe

SUFFIX	ANZAHL
-acja	800
-i/ysta	600
-i/yzm	500
-ator	300
-i/yk	100

Die Analyse des Suffixes ´-ka`

Suffix	das diminuierende ´-ka`	das movierende ´-ka`
Quantität	1180	913
Flexion	Nomen, feminin	Nomen, feminin
Basis	Nomen, feminin	Nomen, maskulin (personal)
Morphologie	11 Alternationstypen	5 Alternationstypen
Zyklen	+	-

Die Analyse des Suffixes ´-ka` (2)

● Regeln für das Suffix ´-ka`

1. $F_{N, \text{gen.pl, fem}} + R + -k(a) \rightarrow F_{N, \text{nom.sing, fem, zIII}}$
2. $F_{N, \text{nom.sing, 1}} + R + -k(a) \rightarrow F_{N, \text{nom.sing, fem, zIII}}$
3. $F_{V, \text{3.sing, präs}} + R + -k(a) \rightarrow F_{N, \text{nom.sing, fem, zIII}}$
4. $F_{\text{Adj, nom.sing}} + R + -k(a) \rightarrow F_{N, \text{nom.sing, fem, zIII}}$

Die Analyse des Suffixes ´-ka` (3)

● Die Abbildung der Regeln:

1. $\text{krów}_{(N, \text{gen.pl, fem})} + R + -k(a) \rightarrow \text{krówk}(a)$
(eine Kuh)
2. $\text{garncarz}_{(N, \text{nom.sing, 1})} + R + -k(a) \rightarrow$
 $\text{garncark}(a)$ (ein Töpfer)
3. $\text{przesył}_{(V, \text{3p.sing, präs})} + R + -k(a) \rightarrow$
 $\text{przesyłk}(a)$ (schicken \rightarrow ein Päckchen)
4. $\text{kiszon}_{(Adj, \text{nom.sing})} + R + -k(a) \rightarrow$
 $\text{kiszonk}(a)$ (eingelegt)

Die Analyse des Suffixes ´-ka` (4)

● Beschreibung der Alternationen

1. $(kwo)k_{N, g.pl, fem} + -k(a) \Rightarrow (kwo)czk(a)$
(eine Henne)
2. $(ko)ść_{N, g.pl, fem} + -k(a) \Rightarrow (ko)stk(a)$
(ein Knochen)
3. $(żagl)ow_{Adj, nom.sing} + -k(a) \Rightarrow (żagl)ówk(a)$
(Segel- → ein Segelboot)

Probleme

- Identifikation der Basis als Derivat

matka (*eine Mutter*) \neq matk(a) \leftarrow mat(a) (*eine Matte*)

- die falsche Richtung der Derivation

książka \Rightarrow książk(a) \rightarrow ksiąg(a) (*ein Band*)

- eine phonologische Umgebung des Suffixes

mistrz(\emptyset) \rightarrow mistrz-yn(i) statt mistrz + -k(a)
(*ein Meister* \rightarrow *eine Meisterin*)

Zusammenfassung

● Perspektiven

- die Beschreibung der Konversion,
- die Sammlung der Präfixe,
- die Entwicklung der formalen Regeln.

● Schlussfolgerung

Die Flexion erweist sich als eine der wichtigsten Komponenten von Wortbildungsprozessen.

Herzlichen Dank für Ihre
Aufmerksamkeit