Introduction
ooo

Linguistic tests
oooooooooooooooooo

Formal derivation and productivity
ooooo

Conclusions
oo

Linguistic tests of productivity for Polish

Joanna Rabiega-Wiśniewska

Institute of Computer Science, Polish Academy of Sciences
jrabiega@ipipan.waw.pl

ATALA Workshop on Morphological Productivity
10th November 2007

Introduction
○○○

Linguistic tests
○○○○○○○○○○○○○○○○○

Formal derivation and productivity
○○○○○

Conclusions
○○

Plan of the Talk

- Introduction – a formal description of word formation
- Motivation
- Tests of morphological productivity
- Formal derivation and productivity
- Conclusions

Introduction

Formal model of Polish nominal derivation:

- Polish word formation description in the framework of computational linguistics
- design of a formalism operating at the grapheme level of the text
- design of an automatic morphological analyzer (synthesizer)

Empirical data:

1. suffixation: *motorówka*, *składak*, *światełko*
2. prefixation: *niepisanie*, *antymateria*, *aliteracja*
3. conversion: *bieg*, *biel*, *geolog*
4. compounding: *lodołamacz*, *wielkolud*

Introduction

Background of formal description:

- list of formatives
- choice of the basis for a given derivative
- inflectional information about the basis and the derivative
- description of alternations
- collection of semantic markers

Motivation

Hoeppner's (1980) proposal of suffixes classification:

- short list of suffixes
- account of both native and foreign suffixes
- suffix variants taken into account
- special terminology excluded
- unproductive and rare formatives excluded

Is productivity of derivational formatives estimated for Polish?

Tests on productivity

There are only three published linguistic tests of morphological Polish productivity.

Common features:

- linguistic questionnaires
- informants among students or academics
- focus on morphological status of given lemmas
- semantics vs. formal structure

Test 1 Aims

The first known test on productivity for Polish was constructed by Alicja Nagórko-Kufel (1977).

Motivation:

- to use linguistic association abilities of average native speakers of Polish
- to state the semantic motivation behind collected derivatives

Test 1 Informants

The test was completed by:

- 212 persons
- science studies students
- high-school students

The author chose students because usually they speak standard Polish. Students of science otherwise than students of humanities have not wide knowledge about language.

Test 1 Questionnaires

The author constructed two tests:

1. Primary questionnaire:
   - a list of derivatives and proposals of their semantic and formal analysis
   - a list of words
2. Final questionnaire:
   - a list of 75 nouns

Test 1 The task

1. The list of derivatives: to choose an appropriate proposal of semantic and formal analysis of derivatives
2. The list of words: to give a lemma that has similar form and meaning or association
2. The list of words: for previous given lemma to state if it interprets the analyzed word

Two groups of 19 students each took the first test. In the first part they rejected mostly words or proposals of definitions that were rare, unusual or they had awaken some emotions. In questionable cases students agreed completely – the author suspected lack of consideration and automatic answering test questions. The final questionnaire consists only of the second list.

Test 1 Results

- The analysis of derivatives was easy if the user could divide/find a derivational morph/formative.

- Formal structure of a derivative was of great importance while choosing its basis comparing to semantic relation between basis candidates.

- Informants chose the simplest bases in structure even if they could give formally more complex but semantically closer one.

- The more complicated structure of derivatives the bigger sets of their definitions, bases, even derivational nests were given.

- There were usually given bases that do not have any alternation comparing to derivatives.

Introduction
ooo

Linguistic tests
ooooo●ooooooooooo

Formal derivation and productivity
ooooo

Conclusions
oo

Test 1 Examples

księgowy ⟵ *księga*
instead of
księgowy ⟵ *księgowość* or *księgować*

'accountant', 'bookkeeper' ← 'book'
instead of
← 'accountancy', 'bookkeeping' or 'to book'

literatura ⟵ *litera* (65%) and *literat* (22%)

'literature' ← 'letter' (65%) and 'literary man', 'writer' (22%)

Test 2

Elżbieta Górska (1982) constructed a test modifying one proposed by Mark Aronoff (1980).

Informants:

- 50 students of two secondary schools

Questionnaires:

- two lists of 38 deverbal nouns cited in the *Index a tergo* (1986)
- first list included nouns as *składacz* 'type-setter' the second, nouns as *składak* 'folding canoe' or 'folding bike'
- the verbal basis in each list were the same, here: *składać* 'to set together, to fold'

Test 2 The task

To decide if a word is:

1. actual FOR A SPEAKER: the speaker uses that word and know what it means

2. possible though non-occurring ACCORDING TO A SPEAKER, i.e. the speaker has never yet used or heard a word but could use the particular word to name sth or sb

3. impossible

Test 2 Verification

Two ways of checking if the test was completed randomly:

- 'control words' i.e. nouns used very often in everyday speech,
  otwieracz 'opener' and *zmywak* 'wash clout'
  if considered as 2 'possible' or 3 'impossible', the test was
  rejected

- two possible words were added to both lists
  *suwacz ← *suwać* 'to push' (*suwak* 'slider, zip')
  *zderzacz ← *zderzać* 'to collide' (*zderzak* 'bumper')
  (*przedłużacz* 'extension rod') **przedłużak* ← *przedłużać* 'to
  lengthen'
  (*otwieracz* 'opener') **otwierak* ← *otwierać* 'to open'
  if marked 1 'actual', the test was considered invalid

Test 2 Assumptions

- Every speaker has their individual lexicon.
- There will be differences between decisions concerning the same words.
- Comparing two derivational formatives that derive the same type of bases would show which one is more productive (more derivates in group 2 – 'potentially exist').

Introduction
○○○

Linguistic tests
○○○○○○○●○○○●○○○○○○

Formal derivation and productivity
○○○○○

Conclusions
○○

Test 2 Results

	SKŁAD**ACZ**	SKŁAD**AK**
1	34%	34%
2	33%	25%
3	33%	41%

- Different treatment of existing words in a language reflects productivity of word formation rules.
- Native speaker of a language can form and accept new words by applying them.

Test 3

Krystyna Kleszczowa (2001) wanted to examine an ability of a speaker to apply derivational rules.

Aim of work:

- how intuition and competence of a user is realized?
- does individual occasional derivatives tell sth about productivity?

Informants:

- academics

Introduction
000

Linguistic tests
00000000000●0000

Formal derivation and productivity
00000

Conclusions
00

Test 3 Questionnaires

A source of derivatives:

- new Polish vocabulary and neologisms (Smółkowa, 1998)

Questionnaire:

- a list of 42 semantic definitions

The task:

- to give an adequate derivative to the given definition

Test 3 Example of definition

'person in the state of depression'

depres**yjnik**	4
depres**iarz**	3
*depres**ant***	3
depresj**onista**	4
depres**tyk**	2
depres**ta**	1
depresj**owiec**	3
lexicon:	*depresj**ant***

Test 3 Examples of definition

'someone who provide an auction'

aukcjarz	
aukciarz	
aukcjonarz	
aukcjonista	
aukcyjnik	
lexicon:	*aukcjoner*

'someone who sells citrus fruits'

cytrusiarz ‖ cytrusarz

Test 3 Results

- Derivatives given by informants usually do not match with those collected in the dictionary.
- Informants used a wide range of word formation formatives, it was difficult to find the most preferred.
- Informants differed even in the way of binding bases and formatives — different alternations appeared in the experiment.
- There is now tendency to chose international or foreign suffixes, especially among younger academics.

Introduction
○○○

Linguistic tests
○○○○○○○○○○○○○○○○●

Formal derivation and productivity
○○○○○

Conclusions
○○

Results of tests

Tests showed that:

- the speaker uses derivational rules (on the flow) but their choice is irregular
- the speaker uses often less complicated rules
- the choice of word formation rule is most influenced by lexical factor

Tests did not show:

- which derivational formatives are productive in contemporary Polish
- and therefore, which of them should be picked for a formal analysis

Formal description of nominal derivation

Criteria for formatives selection:

- a list of formatives extracted from traditional word formation literature
- frequency data from two dictionaries of contemporary Polish (120 and 45 thousands entries)
- almost 70 thousands derivatives considered

Formal description includes:

- 147 formatives for nouns and adjectives
- 392 derivational rules
- 238 alternation rules

Testing derivational rules

To verify the rules a simple test was designed.

Data and tool:

- a list of derivatives from two dictionaries
- a control list of derivative-forms from the PWN corpus
- an inflectional analyzer AMOR (Rabiega and Rudolf, 2001)

Testing derivational rules

Test:

1. analysis of derivatives with AMOR – result: the rate of unrecognized words

2. application of the derivational rule to the derivatives – result: a list of bases

3. analysis of bases with AMOR – result: the rate of unrecognized words

4. comparison of results

Example -K(a)

One of rules:

```
Sr 1.
-K(a)1 : FN.nom.sing.m1 + -K(a)1 → FN.nom.sing.fem
```

Dictionaries	SJPDor	ISJP
Derivatives	2425	579
AMOR analysis		
1st step, unrecognized derivatives	59%	13%
2nd step, unrecognized bases	17%	5%

Future work

The formal model of Polish nominal derivation

- operates on the morphological roots
- is capable of solving a morphological alternation at the boundary of the stem and the affix

How to estimate productivity?

- application of the formal model of derivation
- construction of a morphological analyzer
- tests on Polish large corpora of texts

Conclusions

1. The tests of productivity for Polish have been presented.
2. Their results show rather native speaker intuition, language competence, the way a speaker apply word formation rules.
3. Morphological productivity in texts still remains to estimate
4. The proposal of formal description of Polish derivation and its application has been proposed.

Introduction
000

Linguistic tests
00000000000000000

Formal derivation and productivity
00000

Conclusions
○●

Thank you!