Pertsovas model of learning inflection

Eva Zimmermann
Workshop of the Forschergruppe
'Grammatik und Verarbeitung verbaler Argumente' (University of Leipzig)
Leucorea Wittenberg, 31.08.-01.09.2009

Claim Restrictions about patterns of homonymy in inflectional paradigms reflect something about the way, people learn languages.

Typology	The distribution of different patterns of form-identity is restricted in the languages of the world.
Formal theory	Other generatively equivalent ways to describe syncretisms are inferior to descriptions that use blocking (systematic – accidental syncretism).

$\Rightarrow$ Implementation of a formal learner that is biased to learn no/'simpler' form identity patterns.

Learning Inflection

Assumptions

- morphemes (stems, affixes) stored in the lexicon: form-meaning pairs
- economy: underspecified markers, ø-morphemes, blocking rules

Cross-situational inference

Observing what properties remain unchanged across different situations in which the same form is used.

(1) *Invariant Features*
of affix x (I(x)) is the feature set obtained by intersecting all enviroments in the block of x. Whereas a block for affix x is the set of affix cells in which it occurs.

Form Identity

As a big challenge for cross-situational inference: overgeneralizations.

(2) *Weak German verbal inflection*

		Prs	Pst
Sg.	1.	spiel-e	spiel-t-e
	2.	spiel-st	spiel-t-est
	3.	spiel-t	spiel-t-e
Pl	1.	spiel-en	spiel-t-en
	2.	spiel-t	spiel-t-et
	3.	spiel-en	spiel-t-en

Invariant feature for -e: SG

but ‘there are many other singular contexts, in which the morpheme -*e* does not occur’.

Three types of form-identity

Homonymy The semantic distribution of a morph cannot be described with a single set of necessary and sufficient features.

(3) *homonym 'are' in English*

	SG	PL
1	am	are
2	are	are
3	is	are

(4) *No homonym 'are' in English'*

	SG	PL
1	am	are
2	is	are
3	is	are

1. Natural class syncretism

A (semantic) contrast is neutralized in some sub-paradigms of the grammar.

A	B
A	C

2. Elsewehere homonymy

Cases that can be described with defaults.

A	B
A	A

An underspecified morpheme is blocked in certain contexts by explicit rules specified for certain slots:

(5) Blocking rule: (m, n)
morpheme m blocks morpheme n

3. Overlapping distribution:

Morph x and morph y are in an overlapping distribution if:

1. the invariant features of x and y are consistent with each other, and
2. x occurs in the domain of the invariant features of y and vice versa.

A	B
B	A

Example: German

-Group	+Group	
-e/-ø	-en	+Part, +Speaker
-st	-t	+Part, -Speaker
-t	-en	-Part, -Speaker

invariant feature for -en: [+group]
invariant features for -t: [-speaker]

⇒ consistent with each other
⇒ -en occurs in [-speaker] context and -t in [+group] context

Distribution of form-identity patterns: Typological reality vs. chance frequencies

Computing chance frequencies

features n	paradigm cells	paradigms without homonymy	paradigms with elsewhere h.	paradigms with overlapping h.
1	2	100%	0%	0%
2	4	53%	6	41%
3	8	3%	64%	33%

⇒ if affixes were distributed in a completely random way, paradigms without homonymy would be quite rare.

Paradigms for subject agreement in 30 languages, i.e. 93 paradigms:

	number of paradigms
no form identity	7
only natural class syncretism	41
only elsewhere homonymy	19
only overlapping homonymy	5
mixed patterns	21

⇒ no homonymy in 52 % of the paradigms
⇒ 10% overlapping patterns (no paradigm involved more than one)

Still an underestimation

From the 197 languages in the WALS:

- languages without form-identity were excluded (80)
- languages without verbal agreement for the subject were excluded (57)

⇒ 70 % of the languages were excluded and therefore 85 % of agreement paradigms contain no homonymy at all.

Chance frequency vs. tyopological distribution

- natural class syncretisms (and total irrelevance of features) are more common than homonymy
- homonymy patterns that can be described as defaults are more common than overlapping homonymy patterns

Learning Inflection:

a generalizing, bottom-up learner with a bias for paradigms without homonymy and a strong tendency to avoid overlapping patterns

(idealised) Assumptions

- the learner analyses the semantic information of a given context correctly
- affixes have fixed positions: the learner learns sublexica for different slots
- meaning of inflectional morphemes can exhaustively be described by a combination of some universally given features

(6) *e.g. features for verbal agreement*

[± participant]	1./2. vs. 3.
[± speaker]	1. vs. 2./3.
[± anim]	fem./masc. vs. neut.
[± fem]	fem. vs. masc. (depends on [±anim])
[± group]	Sg vs. Du./Pl.
[± min]	Du. vs. Pl. (depends on value for [±group])

Pertsovas learners...

Three learners that operate within increasingly larger hypotheses spaces:

1. no homonymy
(1:1 form-meaning mapping or natural class syncretism)

2. + elsewhere homonymy

3. + overlapping homonymy

...are one.

One algorithm for learning form-meaning mappings in inflection with two 'check-points' where the algorithm detects instances of homonymy.

Input: pairs (m, e), i.e. text t_j for language L_j
Output: updated lexicon (set of sub-lexicons for specific slots)

The algorithm:

- adds a lexical entry or
- modifies the meaning of an already existing entry (features are removed) or
- adds blocking rules.

Cross-situational learner

Hypothesizing (s, e)

1.) a new lexical entry or
2.) an already existing lexical entry with actualized meaning (intersection of meanings)

Elsewhere learner

⟶ **Is there an elsewhere homonymy?**

Yes!

– hypothesize a blocking rule

No.

– Add (s, e) to Lex_j

General homonymy learner

⟶ **Is there an overlapping homonymy?**

Yes!

– New hypothesis: fully specified input morpheme and new run through algorithm

No.

– Add (s, e) and blocking rule to Lex_j

Complexity

All patterns of form-identity are possible, but some pattern require more time and resurces to learn.

Complexity is measured as a function of:

- number of runs through the algorithm
- size of the resulting grammar: number of lexical items and blocking rules

Quantifying Complexity

learning form identity pattern:

learner 1

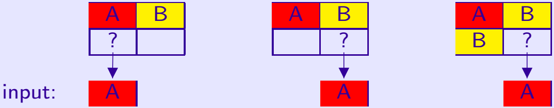
natural class syncretism

learner 2

elsewhere homonymy

learner 3

overlapping homonymy



Learner 1:

- new meaning for A

Learner 2:

- new meaning for A
- BR (B, A)

Learner 3:

- new entry for (A, 2)
- BR (A, B)
- two runs through the algorithm

The learner is biased to learn certain patterns

→ this explains typological asymmetries.

analytical bias: cognitive predisposition making learners more receptive to some patterns (most researchers take this to be UG, but may also emerge from cognitive bases that are not specifically linguistic).

vs. **channel bias** in phonology: some systematic phonetically errors in transmission between speaker and hearer.

Cross-situational learner

Hypothesize 'new' morpheme (m, e):

For all (m, e) in t_j
1. if $\exists$(m, f) $\in$ Lex, then replace (m, f) with (m, f $\cap$ e) in Lex
2. else add (m, e) to Lex.

	+F	-F
+G	A	B
-G	A	C

lexicon: (A) [+F, +G]

input: (A) [+F, -G]
output: (A) [+F]

Elsewhere homonymy

An underspecified morpheme is blocked in certain contexts by explicit rules specified for certain slots:

(7) Blocking rule: (m, n)
morpheme m blocks morpheme n

If two morphemes (m, e) and (m', e') are in the lexicon and e is consistent with e', then:

1. one blocks the other (subset principle: the more specific one blocks the other)
2. or a third morpheme blocks both competitors

Detecting elsewhere homonymy

Hypothesized morpheme (m, e)

Check the lexicon for competitors, i.e. for morphemes whose meaning is consistent with the meaning of (m, e)

If there are competitors:

- Check whether one morpheme is more specific (=it blocks the other)
- Search through your memory whether another morpheme was ever observed in the enviroment that is consistent with the meaning of (m, e) and its competitor (=this will block both)
- Else: the competition remains unresolved
(overgeneralization is predicted until disambiguating data is uncovered)

Detecting elsewehere homomymy

	+F	-F
+G	A	B
-G	A	A

lexicon: (A) [+F]
(B) [-F, +G]

input: (A) [-F, -G]

intersecting meanings (A) [ø]

competitors? (B) [-F, +G]

blocking relation? BR: (B, A)

output: (A) [ø]
(B) [-F, +G]
BR: (B, A)

Overlapping homonymy

Homonymous markers are assumed as last resort.
(phonologically identical morphs are paired with different integers)

Detecting overlapping homonymy

If (m, e) has competitor (s, f):
check whether morphemes with form m as well as form s occur in the contexts that are consistent with the meanings e and f.

i.e.: Set P = set of morphemes that are consistent with the meanings of the currently hypothesized morpheme and all its competitors
If P contains morphemes with form of the currently hypothesized morpheme and all its competitors

	+F	-F
+G	A	B
-G	B	A

Detecting overlapping homonymy

lexicon:	(A) [+F, +G]
	(B) [ø]
input:	(A) [-F, -G]
intersecting meanings	(A) [ø]
competitors?	(B) [ø]
blocking relation?	BR: ?
set P	(A) [ø] and (B) [ø] ⇒ includes forms of both competitors!
new hypothesis:	(A, 2) [-F, -G]
competitors?	(B) [ø]
blocking relation?	BR: (A 1, B)
blocking relation?	BR: (A 2, B)
output:	(A, 1) [+F, +G]
	(A, 2) [-F, -G]
	BR: (A 1, B)
	BR: (A 2, B)
	(B, 1) [ø]

Discussion

- memory stack: learner memorizes everything and searches through all ever heard utterances (to detect blocking relation and overlapping pattern)
- the learner can analyse every context correctly: no errors and no way to go back
- learning bias explains typological reality (diachronic changes?)

Bibliography

Moreton, Elliott (2008). Analytic bias and phonological typology. *Phonology* 25(1):83–127.

Siskind, Jeff (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61(1):39-91, 1996.

Pertsova, Katya (2007). *Learning Form-Meaning Mappings*, PhD thesis, University of California, Los Angeles.

Wilson, Colin (2006). Learning Phonology With Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science*, 30:945-982.