

# Parity and Self-monitoring

Henk Zeevat

CAS, Oslo and ILLC, University of Amsterdam

[henk.zeevat@uva.nl](mailto:henk.zeevat@uva.nl)

Tandem Workshop, Berlin

## Overview

1. Parity

2. Self-monitoring

3. Case studies

I. optional marking

II. np selection

III. freezing

IV. dom and dsm

V. adjectival ordering

VI. additivity

## **Parity**

message is coded into code which is decoded as message1

parity: message=message1

Where is parity in NL?

natural solution:

Grice: speaker intention as message (for any non-natural communication)

Grice's arguments + intersubjectivity + downwards closure

Speech

sequence of words

Alvin Liberman: sequence of articulation gestures

(motor theory of speech perception)

## **Self-monitoring**

two related starting points:

- a. underdetermination of meaning by form
- b. the best hearer's choice is the most probable interpretation in the context

underdetermination of meaning by form

a. argument from context dependency

He did.

John ate a piece of the cake, at yesterday's party.

Kjell-Johan mentioned the papers on antipresupposition in his talk about the milk problem.

Oops.

Look, I made the computer screen go black.

I dropped the scissors.

massive need of contextual integration to get at the message content

Is the context part of the signal?

b. argument from computational linguistics, in particular the collapse of classical parsing

parsing: get correct labelled trees from a sentence

classical: use a grammar

Need large grammars to get full coverage.

Large grammars have more rules

the number of rules increases ambiguity.

Early nineties: near full coverage, 20 word sentence gets 10000 readings in 1.7 seconds.

Result: probabilistic parsing

And that is just the beginning: lexical, resolution, rhetorical structure, integration into common ground

c. evolutionary argument: obliteration of distinctions

Natural Language Understanding is about selecting/constructing one of very many meanings allowed by the form

as in vision, a large part of context integration is stochastic



## Conflicting Observation

(Introspection and experience) We are doing fine.

(Artificial Intelligence) The human processors solve this massive disambiguation problem fast, routinely and with overwhelming success. (vision)

(Psychology) Parity on speaker intention is reached standardly in dialogue (though with feedback loops) and even better in controlled communication (e.g. news bulletin)

## **How parity is reached?**

A. Hearer rationality: pick the most probable interpretation in the context

Going for any other interpretation just increases the chance that the hearer gets it wrong.

Possible counterargument: hearer cannot do this.

Rejoinder: no parity

How to do this is a severely non-trivial question. It is not clear current probabilistic models give the basis for a cognitive theory (too much data, too much number crunching) and that they can be successfully generalised to the higher levels of interpretation. But one should assume that hearers do manage.

cognitive theory: emulation of Bayesian interpretation, centrally using simulated production  
(not in today's subject)

## B. Consequence for the speaker: self-monitoring

Parity will not be reached unless the speaker makes sure the most probable interpretation of her utterance is the interpretation she intended.

correct form for a meaning (syntax, semantics, pragmatics) is not enough

conditioning on probability maximation in interpretation

conditioning on simulated understanding

correct form is the form for which the intended interpretation is the most probable one in the context

## Examples

John and Bill met. He wore a grey coat.

correct pronoun, but parity not guaranteed (he → john, the first)

*sell(a14, a66)*

Bill sold a blue sweater.

choice between: Bill/he/Bill, the new employee/the new employee/an employee/ somebody  
needed to converge on the seller a14.

**Thesis: Parity should be the central problem of linguistics.**

But all linguistic theories seem to be Aristotelian:

grammar is the definition of the relation between forms and meanings

This includes production OT or bidirectional OT.

Such theories do not help with parity.

As accounts of parity: predicted probability of parity is low

though better than random choice of utterance and interpretation

Stochastic CL: only helps in interpretation direction

without self-monitoring:

parity is still a rare event, but probability is up over Aristotelian grammar

Proposal for a production grammar, with self-monitoring using stochastic interpretation

A. Minimal OT syntax (or equivalent)

word order and morphology

alignment constraints, max constraints

B. Self-monitoring component:

second optimisation round

described by a partially ordered set of semantic features

## Semantic Features

Linguistically relevant

Semantically interpretable

Important for communication

Examples:

theta: agent, theme, instrument, ...

number: singular, dual, plural

natural gender: male, female, neuter

rhetorical relation: explanation, elaboration, result, narration. ...

topic question: old, new

correction: yes, no

constraint interpretation of a semantic monitoring feature  $f$

constraint  $mon(f)$ , monitoring for  $f$

(complication about multiple instantiations of  $F$  in  $I$  ignored here)

$I^f$ : the value that the interpretation  $I$  assigns to  $f$

$mon(f)$  assigns an error to  $U$  under  $I$  iff

$\exists J(p(I|U) \sim p(J|U) \wedge J^f \neq I^f)$

$mon(F)$  gives an error to  $U$  for  $I$  iff there are roughly as probable or more probable alternative inputs  $J$  for which  $U$  is optimal with  $J$  diverging from  $I$  in the value of  $f$

Marking

no errors on  $mon(f)$ :  $U$  **marks**  $f$

errors on  $mon(f)$  on a winner  $U$ :  $C$  does not mark  $f$  but there is no better alternative



## **typical profile of a self-monitoring application**

soft edge/exceptions

no ungrammaticality but an unintended change to the intended meaning

feature controls the phenomenon

the enforced marking must be overt and is not tied to a particular marking device

syntax: Russian nouns maximise case

but some words (mat'(nom or acc), doc'(nom or acc), kofe (nom or acc)) exhibit case

syncretism and so do not mark their theta-role

## Applications of Self-monitoring I

### **Optional Discourse Markers**

John fell. Bill pushed him.

John fell. Then Bill pushed him.

John fell. Although Bill pushed him.

John fell. Mary smiled at him.

John fell. Because Mary smiled at him.

John fell. Although Mary smiled at him.

To describe:

If the intended discourse relation is not the default a hearer would infer, it must be marked by an overt marker, otherwise marking can happen but is less preferred

pure syntax attempt

assumption: RR is included in input

max(RR): mark the discourse relation

problem: discourse relations are mostly not overtly marked

Monitoring account

RR is marked in input

mon(RR)

causes the marked form to win when the interpretation is not the default in the context and the unmarked form to be better when the input value is the clear default

the treatment can be generalised to other optional marking

e.g. optional progressive marking in Dutch, German and Norwegian (but not French or English where it is syntax)

past tense in Chinese

definiteness in Russian

assumption for such applications: parity is also reached on these features by speakers of these languages

## Applications of Self-monitoring II

### **Pronouns and Ellipsis**

the referential hierarchy of psychological concepts (Gundel, Hedberg, Zacharski):

IN FOCUS > ACTIVATED > FAMILIAR > UNIQUELY IDENTIFIABLE > REFERENTIAL >  
TYPE IDENTIFIABLE

or an extension and reduction of it

FIRST > SECOND > REFLEXIVE > IN FOCUS > ACTIVATED > FAMILIAR > UNIQUELY  
IDENTIFIABLE > REFERENTIAL > TYPE IDENTIFIABLE

is roughly aligned with a (markedness?) hierarchy of classes of forms e.g.

∅, first > second > reflexive > pronoun > pronoun, demonstrative > name, demonstrative,  
definite > indefinite demonstrative > indefinite, bare

## Alignment

1. very imperfect (Gundel, Hedberg & Zacharski)
2. inventory dependent (pronouns and agreement)
3. syntax dependent (ellipsis)
4. pragmatics (suffice for identification)

My guru and his disciple. (Isherwood)

Your humble servant.

Everybody voted for John . Even John voted for John.

(In the mirror). I like you/me/myself/him.

John and Bill came to visit. John/ \*he ....

A waiter/the grey haired waiter/the guy who you met last year at the kindergarden explained the menu.

A monitoring hierarchy

ID > POLITE > FIRST > SECOND > REFLEXIVE > RECIPROCAL > IN FOCUS >  
ACTIVATED > FAMILIAR

ID: discourse referents as values

FIRST: ik, wij

SECOND: jij[fam], u, jullie [fam]

REFLEXIVE: zich, zichzelf

RECIPROCAL: elkaar

IN FOCUS: ie [fam], hij, zij, het

ACTIVATED: die, die N, deze N

FAMILIAR: die N [fam], de N

## Applications of Self-monitoring III

### **Word order freezing**

Mat' ljubit papu/a. Mother loved father/father loves mother.

Mat' ljubit doc'. Mother loves daughter.

difference: doc' (and mat') are the same in nominative and accusative

Quite widely attested: Hindi, Korean, Japanese, Latin, German, Dutch

Typical: case marking or head marking restores word order freedom so does semantic plausibility and possibly even parallelism

Grass eats the horse (in Dutch or German)

Kafe ljubit doc' (a garden path effect is reported)

Who does Peter love?

Maria liebt Peter (nicht Hanna).



The descriptive problem:

predict that a is ambiguous

a. Welches Maedchen liebt Peter? (What girl does Peter love/loves Peter?)

while letting b and c be unambiguous

b Ihn liebt Maria. (Maria loves him)

c. Maria liebt Peter. (Maria loves Peter)

production syntax for German

WH-FIRST > PROMINENCE

How to do freezing?

Production restrictions will have to look at lots of (overt) factors (case marking, headmarking, selection restrictions, context) which makes them lack universality

theta monitoring: assign marks for probable theta-variant interpretations

$love(p, m), love(m, p)$

Maria liebt Peter<sub>obj</sub> vs Peter<sub>obj</sub> liebt Maria

(given prominence) the interpretation:  $love(p, m)$  is more probable for “Peter liebt Maria” but not for “Maria liebt Peter”

WH-FIRST uses up the ordering possibilities, so no competing form

## Applications of Self-monitoring IV

### **DOM and DSM** (sketch)

optional case marking, obligatory passivisation, inverse marking ...

Silverstein's generalisation

universal probabilities

indefinite subjects are rare

inanimate subjects are rare

animate objects are rare

me and you are a bad objects

...

marking devices

accusative marking: marks theta

ergative marking: marks theta

passive: marks theta by making subject oblique

inverse: marks theta

freezing: can mark theta by word order

verb agreement: can mark theta

typology

marker inventory

syntax (obligatory marking)

evolution

what happens given syntax plus marker inventory plus theta monitoring

events:

recruiting (oblique marker to ergative marker, preposition to accusative case marker, article to case marker)

reinforced marking makes marking more necessary or even obligatory

slow increase of making under monitoring (overmarking does not get punished, so other factors have to overcome ever stronger bias))

needed constraints: same as for freezing

example: slow progress of Spanish object marking (Aissen)

evidence: obligatory case marking/passivisation/inverse marking/word order (Aissen) without functional necessity

## Applications of Self-monitoring V

I have an Italian tallest student.

Inherence ordering:

kind > material > shape > color > size > number

3 big red square wooden houses

mon(scope,adj)

mon(inherence, adj)

mon(scope) > mon(inherence)

issue: mon(F) intermingled or below hard constraints

other argument: hearer oriented max constraints (Boersma)



## Applications of Self-monitoring VI

### Additivity

*too* is obligatory (Green, Kaplan)

explanation by  $\max(X)$  (Singh:presupposition, Zeevat:other)

prediction: *too* is always obligatory

clearly false:

1. not needed in list answers 2. omitting *too* in corpus

The English-Norwegian Parallel Corpus

Two probes on “too”

24 omissions lead to anomaly, 5 not clearly so, 1 not.

31 omissions lead to anomaly, 14 not clearly so, 6 not.

(typically cannot see cases where *too* is omitted)

Gregoire Winterstein's thesis: sometimes *too* is not allowed even though the presupposition is common ground

John did not solve all the problems and George solved some problems too.

Self-monitoring can be applied to “obligatory” marking of additivity

probability: topic questions are mostly new

additivity (too) and replacement (instead): a topic question is readdressed

mon(TQ): old, new

mon(CORRECTION): yes, no

default: TQ = new (and therefore CORRECTION = no)

additive: TQ=old, CORRECTION = no

replacement: TQ = old, CORRECTION = yes

exceptions to obligatory additive marking:

list answers (clear from context that additivity is intended)

the question has not been a topic question properly, but only implicitly (can be left that way)

accommodation cases (TQ is discourse new, the old answer needs to be guessed and there is no need for doing this)

Winterstein: no antecedent addressing of topic question

(What did the boys have?)

John had spaghetti, Bill had spaghetti and Tim had pizza.

John went to McDonalds. Bill had a bad meal too.

Do you want some coffee too?

John didn't get some of the answers. ???Bill got some answers too.

additivity is worth monitoring for

## INTERPRETATION

Do not miss anaphoric opportunities! (Williams)

Don't accommodate! (Blutner)

NEW (myself)

Minimal models (I and II)! (Kathrin Schulz, Hamm and Lambalgen, many others)

$?xQx$

$a$

$?xQx$

$b$

$a = b$  (I),  $b$  instead of  $a$  (II),  $a + b$  (loses from I and II)

## Conclusions

1. ASM gives a different kind of obligation with a soft edge
2. Self-monitoring by simulated stochastic interpretation predicts a direct effect of probabilities on form
3. As observable in the typology of DSM/DOM
4. “Too” is not very different from “because”. The main difference is in the source of the probabilities. There is nothing optional about “because” either in:  
  
John fell because Mary smiled at him.