Generalized Reference
Referring with and without language by matching, pointer, or address

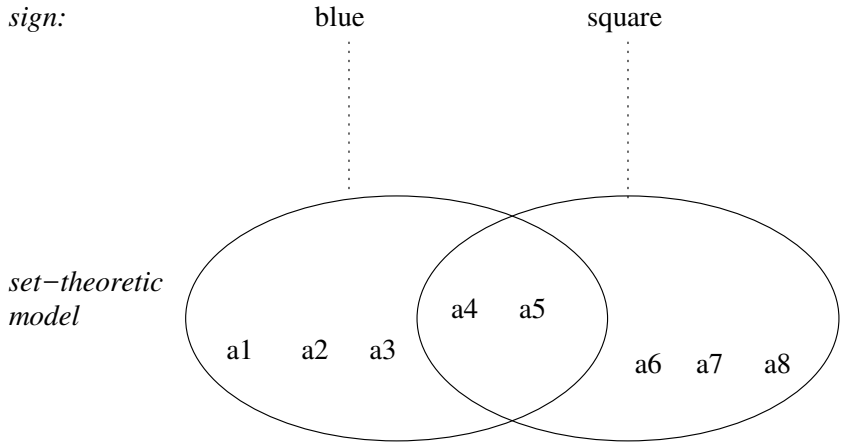
Roland Hausser

July 13, 2017

1 What is Reference?

1.1 Sign-based-based Ontology

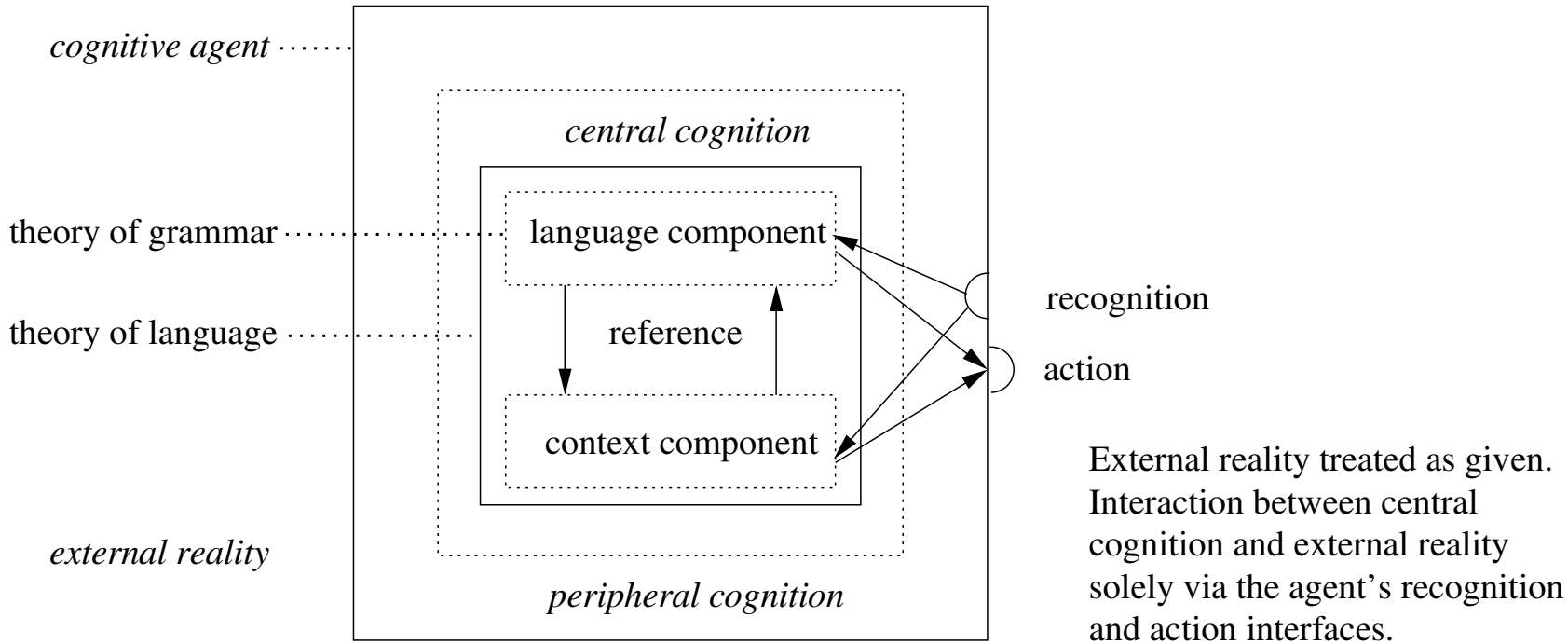
Direct Relation between the language sign and the set-theoretic model



Formally: F(blue) = {a1,a2, a3, a4, a5}
F(square) = {a4, a5, a6, a7, a8}

No agent in an environment
No interfaces for recognition or action
No database (memory)
No speak mode and no hear mode

1.2 Agent-based Ontology



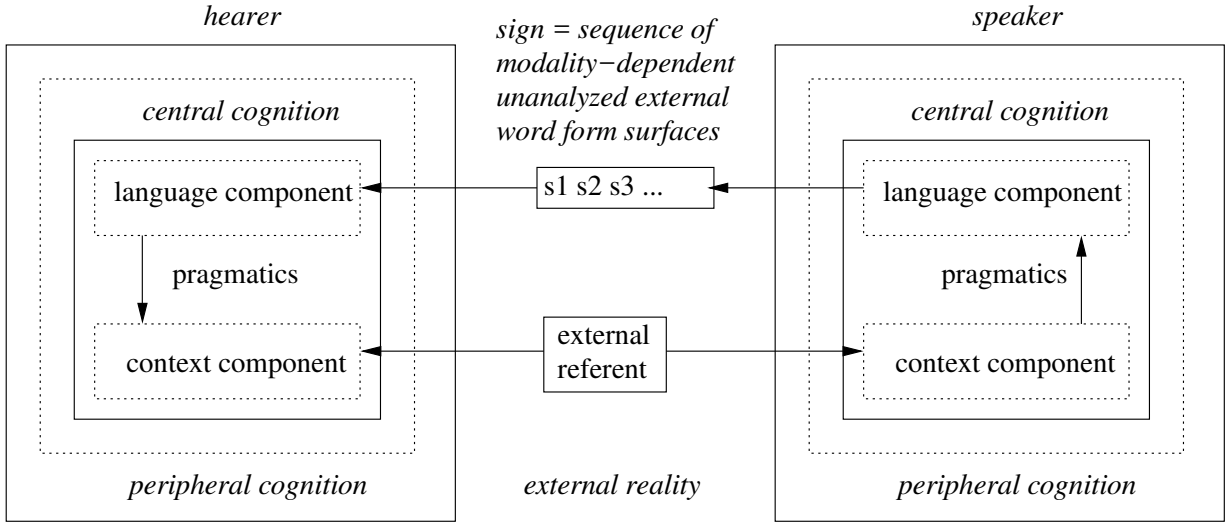
Reference is generalized because it relates *contents*, regardless of whether they are attached to language surfaces or not.
Examples of reference without language:
• Identifying a current nonlanguage recognition with something seen before.
• Identifying a nonlanguage recognition with an earlier language content, e.g. something read (for example, in a guide book) or heard about.

The external surfaces have neither meaning nor any grammatical properties, but may be measured by the Natural Sciences. In cognition, contents are attached to language-dependent surfaces by means of *conventions* which every member of the community has to learn.

1.3 Immediate vs. mediated reference

- *Immediate reference* is the speaker's or hearer's reference to objects in the current task environment.
- *Mediated reference* is the speaker's or hearer's reference to objects which exist in memory, but not in the current task environment, e.g. *Aristotle*.

1.4 Immediate reference as a purely cognitive procedure



There is no external relation between the sign and the referent (no reification!)

2 Type/Token Matching of Elementary Concepts

2.1 Type and token of the color called blue

type

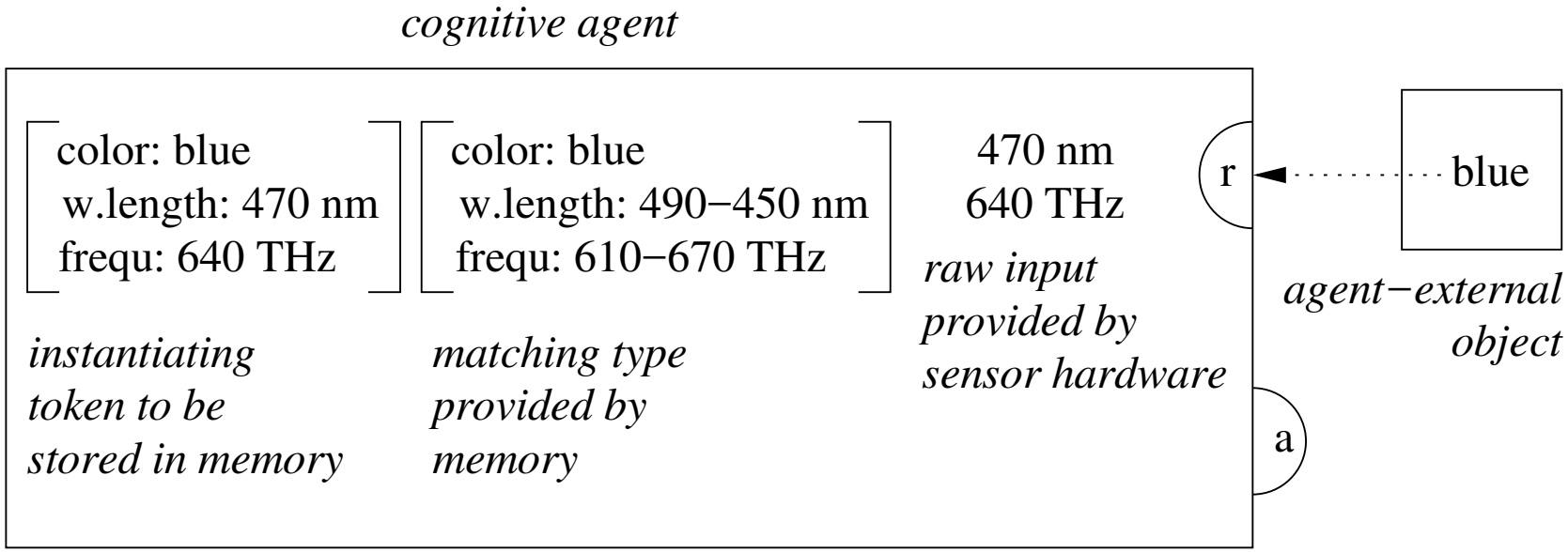
$$\begin{bmatrix}\text{wavelength: } \alpha \\ \text{frequency: } \beta\end{bmatrix}$$

where $\alpha \epsilon$ 490–450 nm
and $\beta \epsilon$ 610–670 THz

token

$$\begin{bmatrix}\text{wavelength: 470 nm} \\ \text{frequency: 640 THz}\end{bmatrix}$$

2.2 Concept type and token in color recognition



2.3 Concept type and token of the concept square

type

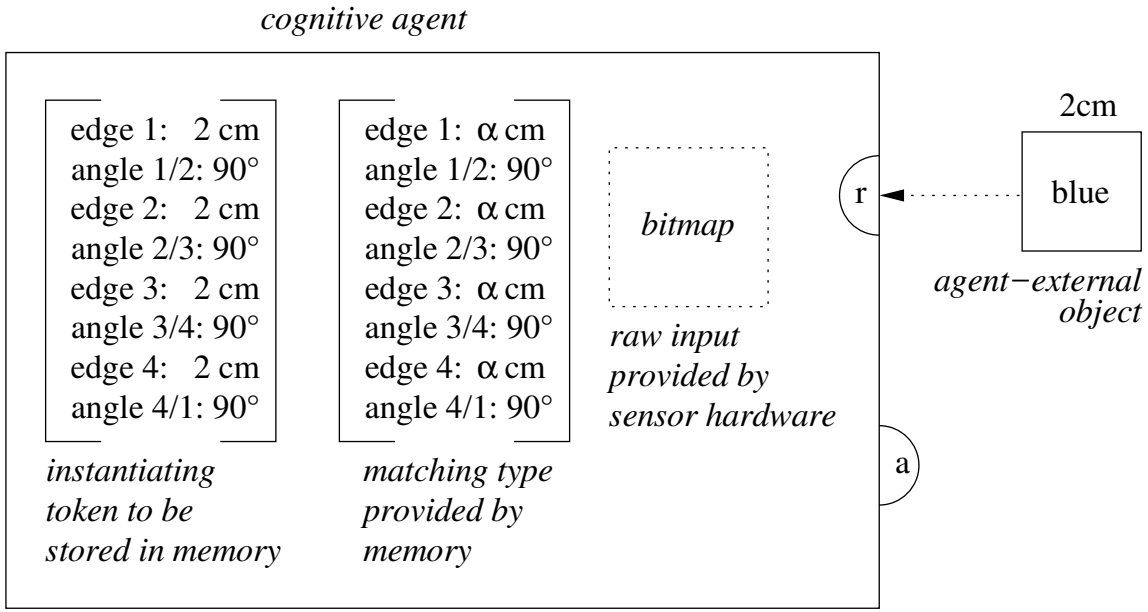
$$\begin{bmatrix} \text{edge 1: } \alpha \text{ cm} \\ \text{angle 1/2: } 90^{o} \\ \text{edge 2: } \alpha \text{ cm} \\ \text{angle 2/3: } 90^{o} \\ \text{edge 3: } \alpha \text{ cm} \\ \text{angle 3/4: } 90^{o} \\ \text{edge 4: } \alpha \text{ cm} \\ \text{angle 4/1: } 90^{o} \end{bmatrix}$$

where α is a length

token

$$\begin{bmatrix} \text{edge 1: 2 cm} \\ \text{angle 1/2: } 90^{o} \\ \text{edge 2: 2 cm} \\ \text{angle 2/3: } 90^{o} \\ \text{edge 3: 2 cm} \\ \text{angle 3/4: } 90^{o} \\ \text{edge 4: 2 cm} \\ \text{angle 4/1: } 90^{o} \end{bmatrix}$$

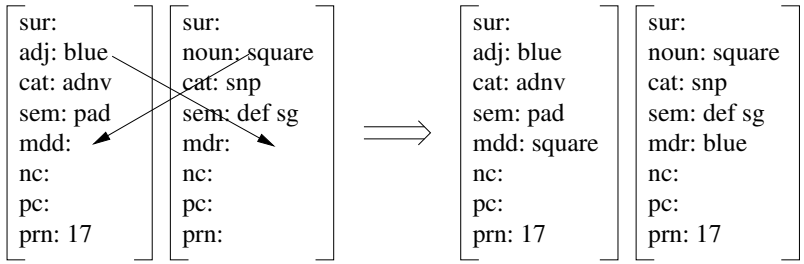
2.4 Type and token in recognizing a square



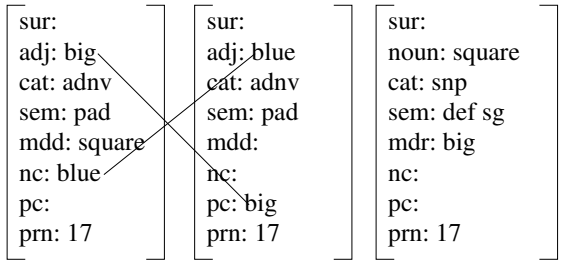
3 Data Structure and Database Schema

For composition, concepts are embedded as *core values* into flat (nonrecursive) feature structures with ordered attributes, called *proplets*. Serving as the computational data structure of DBS, proplets are connected by address, implemented by cross-copying:

3.1 Functor-Argument: concatenating proplets by cross-copying



3.2 Coordination: concatenating proplets by cross-copying



A content is defined as a set of proplets (order-free) which are connected by the semantic relations of structure, i.e. functor-argument and coordination, encoded as proplet-internal *addresses*.

	member proplets	*now front*	*owner values*
	[adj: big cat: adnv		big
...	mdd: square nc: blue prn: 625]		
	[adj: blue cat: adnv		blue
	... mdd: pc: big prn: 625]		
	[noun: square cat: def sg		square
...	fnc: mdr: big prn: 625]		

A word bank has a two-dimensional structure like a classic network database. The horizontal *token lines* consist each of an owner value, preceded by a free slot called the *now front*, preceded by member proplets with the same core value (which equals the owner) in the order of their arrival. Vertically, the token lines are in the alphabetical order induced by their owner values.
Because proplets are order-free, they may be stored as needed by the database (word bank) without losing their semantic relations. The primary key consists of the core and the prn value.

4 *The Three Kinds of Reference

4.1 Reference by type-token matching

occurs with nouns (objects), verbs (relations), and adjectives (properties)

noun: table, chair, dog, square, ...
verb: walk, talk, see, give, ...
adj: blue, fast, big, beautiful,...

4.2 Reference by pointing

occurs with nouns (objects) and adjectives (properties)

noun: pronouns like I, me, you, he, him, she, her, ...
adj: here, now, there, then, ...

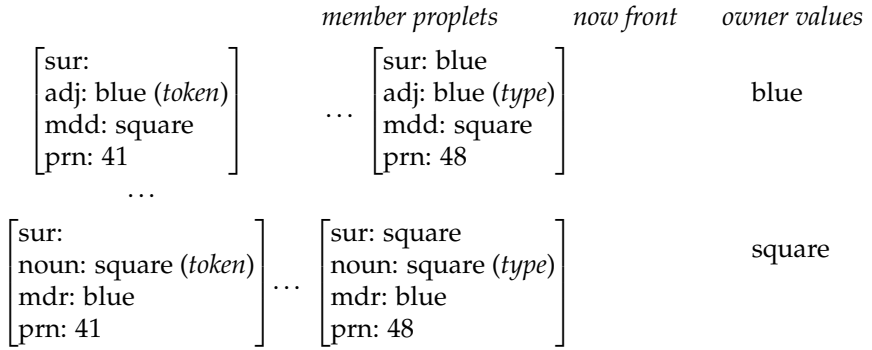
4.3 Reference by address (coreference)

occurs with nouns (objects) only

noun: (dog 23), realized as dog/Fido/he/him
name: (person 14), realized as John, he/him
pronoun: pro3, realized as John, he/him

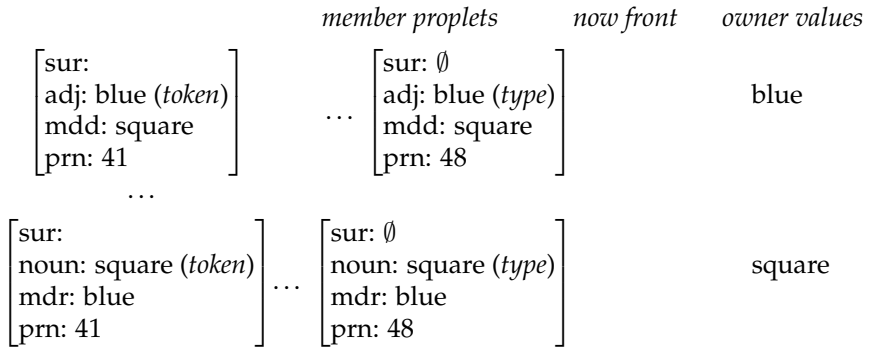
5 Reference by Matching Concept Types and Concept Tokens

5.1 Reference with language proplets in token lines: blue square



The types have non-empty sur slots.
Speaker goes from tokens to types, hearer from types to tokens.
The two types and the two tokens are each connected by address.

5.2 Reference by matching without language



The types have empty sur slots, emphasized here by ∅.
Without surfaces there is no inter-agent communication.
The types serve for classification of the tokens.

5.3 *Constellations of generalized reference with concepts

1. *Nonlanguage content referring to nonlanguage content*
   Example: Agent identifies something seen with something seen before.

2. *Language content referring to nonlanguage content*
   Example: Agent describes a landscape in speak mode.

3. *Nonlanguage content referring to language content*
   Example: Agent identifies a current nonlanguage recognition with something it has read (for example, in a guide book) or heard about before.

4. *Language content referring to language content*
   Example: Agent describes what it has heard or read.

DBS reference is generalized not only because it applies to concept matching with and without language, but also because it includes the referring with pointers and by address.

6 Reference by Pointing (Indexical)

6.1 1ST PERSON PRONOUNS DISTINCTIONS

sur: I	sur: me	sur: we	sur: us	
noun: pro1	noun: pro1	noun: pro1	noun: pro1	
cat: s1	cat: obq	cat: p1	cat: obq	
sem: sg	sem: sg	sem: pl	sem: pl	
fnc:	fnc:	fnc:	mdr:	Same pointer, different cat and sem values (grammatical differentiation).
mdr:	mdr:	mdr:	fnc:	
nc:	nc:	nc:	nc:	
pc:	pc:	pc:	pc:	
prn:	prn:	prn:	prn:	

6.2 REPRESENTING I heard you. AS A LANGUAGE CONTENT

sur: I	sur: heard	sur: you	
noun: pro1	verb: hear	noun: pro2	
cat: s1	cat: #n′ #a′ decl	cat: sp2	
sem: sg	sem: past	sem:	
fnc: hear	arg: pro1 pro2	fnc: hear	Example of a proposition
mdr:	mdr:	mdr:	
nc:	nc:	nc:	
pc:	pc:	pc:	
prn: 71	prn: 71	prn: 71	

6.3 TOKEN LINE EXAMPLE OF STARS DEFINED AS PROPLETS

		member proplets		*now front*	*owner value*	
S: kitchen	S: kitchen	S: living room	S: garden			
T: t_1– t_2	T: t_{2+1}– t_3	T: t_{3+1}– t_4	T: t_{4+1}– t_5			
A: Sylvester	A: Sylvester	A: Sylvester	A: Sylvester		Sylvester	The STAR (i) encodes the agent's onboard orientation and (ii) provides values for indexical pointers.
R: Speedy	R: Speedy	R: Tweety	R: Hector			
3rd:	3rd:	3rd: Speedy	3rd:			
prn: 63–70	prn: 71–78	prn: 79–82	prn: 83–87			

6.4 ANCHORING A CONTENT TO A STAR-0

sur:	sur:	sur:	S: kitchen	The proposition and the STAR are connected by having the same prn value
noun: pro1	verb: hear	noun: pro2	T: 2013-09...	
cat: s1	cat: #n′ #a′ decl	cat: sp2	A: Sylvester	
sem: sg	sem: pres	sem:	R: Speedy	
fnc: hear	arg: pro1 pro2	fnc: hear	3rd:	
mdr:	mdr:	mdr:	prn: 63	
nc:	nc:	nc:		
pc:	pc:	pc:		
prn: 63	prn: 63	prn: 63		

6.5 SPEAK MODE ANCHORING TO A STAR-1

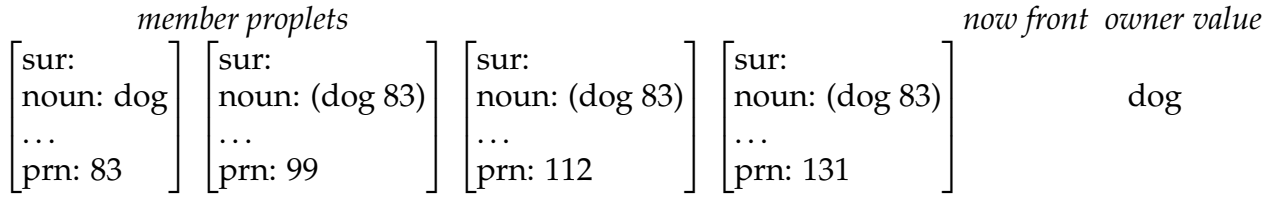
sur: I	sur: heard	sur: you	S: kitchen	The STAR-1 differs from the STAR-0 in the tense of the verb and the prn value
noun: pro1	verb: hear	noun: pro2	T: 2013-09...	
cat: s1	cat: #n′ #a′ decl	cat: sp2	A: Sylvester	
sem: sg	sem: past	sem:	R: Speedy	
fnc: hear	arg: pro1 pro2	fnc: hear	3rd:	
mdr:	mdr:	mdr:	prn: 64	
nc:	nc:	nc:		
pc:	pc:	pc:		
prn: 64	prn: 64	prn: 64		

6.6 STAR-2 PERSPECTIVE IN HEAR MODE

sur:	sur:	sur:	S: kitchen	The STAR-2 perspective converts the speaker's I heard you to You heard me with concomitant adjustment of the STAR values
noun: pro2	verb: hear	noun: pro1	T: 2013-09...	
cat: sp2	cat: #n′ #a′ decl	cat: obl	A: Speedy	
sem:	sem: past	sem: sg	R: Sylvester	
fnc: hear	arg: pro2 pro1	fnc: hear	3rd:	
mdr:	mdr:	mdr:	prn: 53	
nc:	nc:	nc:		
pc:	pc:	pc:		
prn: 53	prn: 53	prn: 53		

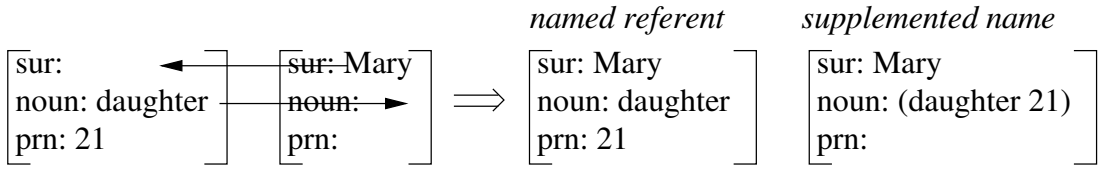
7 Reference by Generalized Baptism (Name)

7.1 Token line showing object permanence by address



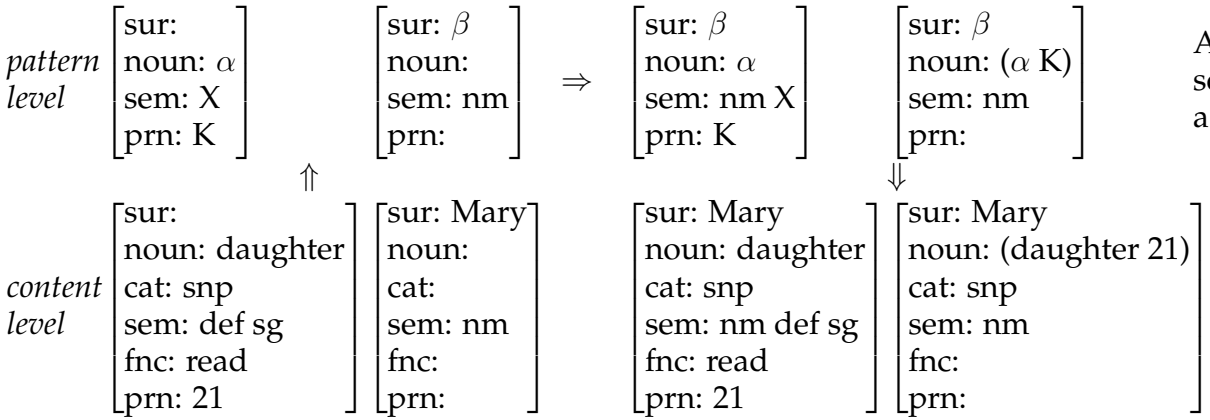
Object-permanence (Piaget) is a non-truthconditional, non-modal alternative to Kripke's rigid designators

7.2 Baptism as cross-copying



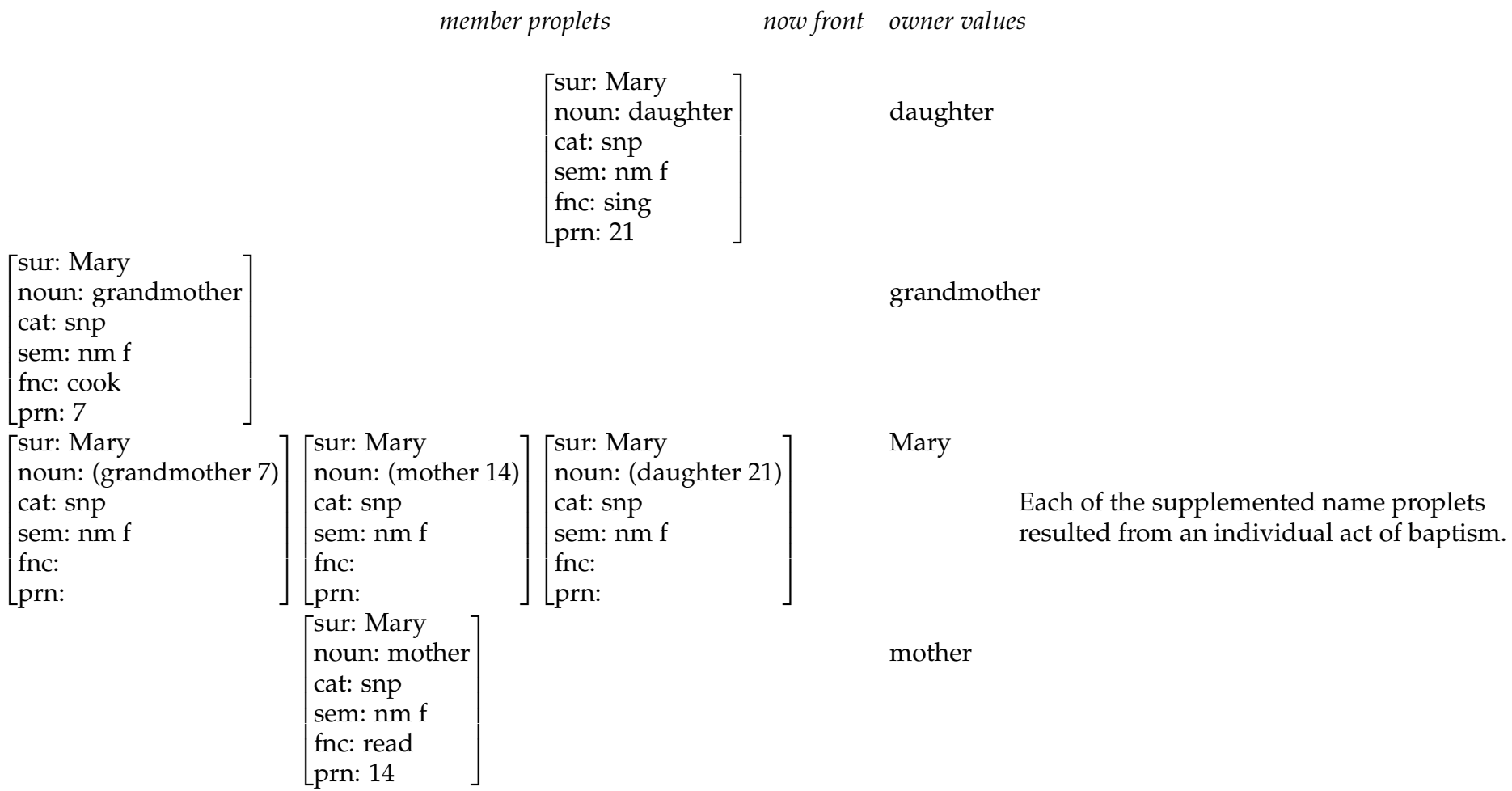
Reference by name is generalized because the baptism may be explicit or implicit Named referent is used in the speak mode, supplemented name in the hear mode.

7.3 Applying the formal baptizing inference



Applies whenever someone or something unknown is called a certain name

7.4 Name referring with multiple referents



8 *Reference by Address (Coreference)

8.1 Coreference in a token line

member records			*now front*	*owner value*
noun: unicorn sem: indef sg prn: 23	…	noun: (unicorn 23) sem: def sg prn: 26		unicorn

8.2 Indexical interpretation of him

noun: Speedy fnc: hide prn: 78	verb: hide arg: Speedy mdr: cupboard nc: (hear 79) prn: 78	noun: cupboard mdd: hide prn: 78		
noun: pro2 fnc: hear prn: 79	verb: hear arg: pro2 pro3 pc: (hide 78) prn: 79	noun: pro3 fnc: hear prn: 79	S: garden T: 2013-09-23T08:20:00 A: Hector R: Sylvester 3rd: Tweety prn: 78-79	

8.3 Coreferential interpretation of him

noun: Speedy fnc: hide prn: 78	verb: hide arg: Speedy mdr: cupboard nc: (hear 79) prn: 78	noun: cupboard mdd: hide prn: 78		
noun: pro2 fnc: hear prn: 79	verb: hear arg: pro2 (Speedy 78) pc: (hide 78) prn: 79	noun: (Speedy 78) fnc: hear prn: 79	S: garden T: 2013-09... A: Hector R: Sylvester 3rd: prn: 78-79	

Further Reading

Hausser, R. (1992) "Complexity in left-associative grammar," *Theoretical Computer Science*, Vol. 106.2:283–308

Hausser, R. (1999/2001/2014) *Foundations of Computational Linguistics, Human–Computer Communication in Natural Language, 3rd ed.*, pp. 522, Springer

Hausser, R. (2006) *A Computational Model of Natural Language Communication – Interpretation, Inference, and Production in Database Semantics*, Springer (preprint 2nd ed. available online at `lagrammar.net`)

Hausser, R. (2011) *Computational Linguistics and Talking Robots – Processing Content in Database Semantics*, Springer (preprint 2nd ed. available online at `lagrammar.net`)

Hausser, R. (2015) "From Montague Grammar to Database Semantics," *Language and Information*, ISSN 1226-7430, Vol. 19(2):1-16

Hausser, R. (2017a) "A computational treatment of generalized reference," *Complex Adaptive Systems Modeling*, Vol. 5(1):1–26. Also available at http://link.springer.com/article/10.1186/s40294-016-0042-7

Hausser, R. (2017b) *How to build a Talking Robot – Linguistics, Philosophy, and Artificial Intelligence*, pp. 120, Springer

Hausser, R. (2017c) *Twentyfour Exercises in Linguistic Analysis. DBS software design for the hear and the speak mode of a talking robot*, 318 pp. Available online at `lagrammar.net`