ICCG5, University of Texas at Austin

An Object-Oriented Computational Model for Processing Linguistic Constructions

Yoichiro Hasebe

Doshisha University

National Institute of Information and Communications Technology, Japan

yohasebe@gmail.com

http://yohasebe.com

Introduction

1. The System
   - Advanced Multi-Layered Data Processing System (AMPS)
2. Background
   - Linguistics and OOD
   - Symbolic nature of language
   - Creating a meta framework of corpus processing
3. Solutions
   - Object-oriented and symbolic-oriented architecture
   - Multi-layered data matrix format
   - Formal Concept Analysis
4. Conclusion
   - Possible Applications

The System

- Overview of AMPS
  1. Accepts a collection of annotated text data and store it as a corpus
  2. Designed in accordance with some of the recent, non-derivational linguistic theories
  3. Detects construction patterns in the corpus and creates a hierarchical network of the patterns
  4. Web-based interface developed on Ruby and Rails that offers various views, such as ones specialized in morphemes, frames, or phrases, as well as constructions.

AMPS

http://localhost:3000/construction/show/2

Google

| manage data | create template | logout |

AMPS
Advanced Multi-Layered Data Processing System

Documentation | Constructions | Frames | Primitives | Morphemes | Phrases

Construction ID:EVCA-Putting02

An instance of construction
(Table representation)

Basic information

- File name: EVCA Putting02.xml
- Morpheme set: **Cheryl stood the books on the shelf**

Table representation

		M1	M2	M3	M4	M5	M6	M7
	Morpheme original pos	Cheryl Cheryl N	stood stand V	the the D	books book N	on on P	the the D	shelf shelf N
F1	PUT	AG	HEAD	TH		ON		
F2	ON					HEAD	LOC	

Frame relationship

- Frame **PUT** might presuppose frame **ON**

AMPS

http://localhost:3000/construction/show_tree/2

Google

Construction ID:EVCA-Putting02

Basic information

- File name: EVCA-Putting02.xml
- Morpheme set: **Cheryl stood the books on the shelf**

Tree representation

Show Table | Show Tree | Download XML | Download DOT

An instance of construction
(Tree representation)

Frame relationship

- Frame **PUT** might presuppose frame **ON**

AMPS
Advanced Multi-Layered Dat

frames

Filter on item name :

Displaying frames **1 - 15** of **24** in tot

« Previous 1 2 Next »

Name	Count
REMOVE	41
SEND/CARRY	34
PUT	32
FROM	21
TO	17
WITH	8
ON	7
OF	6
AROUND	5
ONTO	4
ACROSS	4
AT	3
BEHIND	3
UNDER	3
INTO	2

AMPS
Advanced Multi-Layered Data

primitives

Filter on item name :

Displaying primitives **1 - 15** of **94** in to

« Previous 1 2 3 4 5 6 7

Name	Count
REMOVE:HEAD	41
REMOVE:AG	37
SEND/CARRY:AG	34
SEND/CARRY:HEAD	34
PUT:HEAD	32
SEND/CARRY:TH	29
PUT:AG	27
PUT:TH	25
REMOVE:TH	22
FROM:HEAD	21
FROM:LOC	17
TO:HEAD	17
TO:LOC	16
REMOVE:LOC	15
SEND/CARRY:TO	14

AMPS
Advanced Multi-Layered Data

morphemes

Filter on item name :

Displaying morphemes **1 - 15** of **190** in

« Previous 1 2 3 4 5 6 7

Name	Count
the	142
book	26
from	21
Nora	18
to	16
Carla	14
table	13
Brian	10
wipe	10
of	9
her	8
Jessica	8
clear	8
Amanda	8
on	7

AMPS
Advanced Multi-Layered Data Process

phrases

Filter on item name :

Displaying phrases **1 - 15** of **248** in total

« Previous 1 2 3 4 5 6 7 8 9

Name	Count
from	21
Nora	18
to	16
the book	15
Carla	15
the table	12
the books	9
cleared	9
Brian	9
Jessica	8
with	8
Amanda	8
on	7
her	7
slid	7

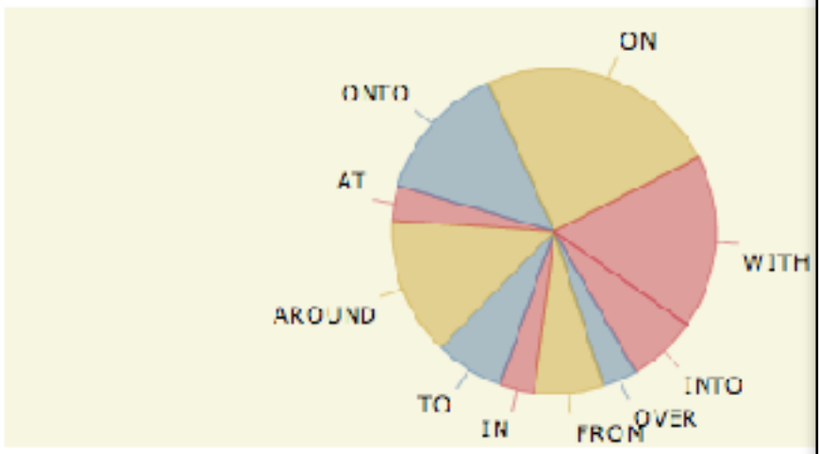
Information on frame PUT

Links to other views

- Constructions containing frame PUT

Frames possibly presupposed by PUT

Name	Count
ON	7
WITH	5
ONTO	4
AROUND	4
TO	2
INTO	2
FROM	2
OVER	1
IN	1
AT	1



Primitives included in PUT

Name	Count
PUT:HEAD	32
PUT:AG	27
PUT:TH	25
PUT:ON	7
PUT:LOC	7
PUT:WITH	5
PUT:ONTO	4
PUT:AROUND	4

Information on morpheme from

Links to other views

- Constructions containing morpheme fr

Instances of morpheme from

Name	Count
from (from, P)	21

Morphemes sharing same attributes

Name	Count
to	16
of	9
with	7
on	7
around	5
onto	4
across	4

Information on primitive REMOVE:HEAD

Links to other views

- Constructions containing primitive REMOVE:HEAD

Phrases that REMOVE:HEAD corresponds to

Name	Count
cleared	9
wiped	6
shoveled	5
was wiping	3
was shoveling	2
stole	2
removed	2
mopped	2
cheated	2
banished	2
trimmed	1
mined	1
is clear	1
deboned	1
cured	1
boned	1

Information on phrase the book

Links to other views

- Constructions containing phrase the book

Primitives corresponding the book

Name	Count
SEND/CARRY:TH	12
PUT:TH	2
SEND/CARRY:TO	1

Primitives neighboring the book

Name	Count
SEND/CARRY:HEAD	13
SEND/CARRY:AG	13
SEND/CARRY:TO	6
SEND/CARRY:LOC	2
SEND/CARRY:FROM	2
PUT:HEAD	2
PUT:AG	2
SEND/CARRY:WITH	1
SEND/CARRY:BNF	1
PUT:ONTO	1
PUT:ON	1

AMPS

http://localhost:3000/s_construction/list

Google

| manage data | create template | logout |

AMPS
Advanced Multi-Layered Data Processing System

Documentation | Constructions | Frames | Primitives | Morphemes | Phrases

Constructions

List of all construction instances

Data | Network (simple) | Network (detailed) | Download DOT

Filter on item name :

Displaying constructions **1 - 15** of **90** in total

« Previous 1 2 3 4 5 6 Next »

Name	ID	C
REMOVE:HEAD	2	4
REMOVE:AG+REMOVE:HEAD	4	3
SEND/CARRY:AG+SEND/CARRY:HEAD	10	3
PUT:HEAD	8	3
SEND/CARRY:AG+SEND/CARRY:HEAD+SEND/CARRY:TH	9	2
PUT:AG+PUT:HEAD	6	2
PUT:HEAD+PUT:TH	12	2
REMOVE:HEAD+REMOVE:TH	16	2
FROM:HEAD	38	2
PUT:AG+PUT:HEAD+PUT:TH	13	2
REMOVE:AG+REMOVE:HEAD+REMOVE:TH	15	2
FROM:HEAD+FROM:LOC	37	1
TO:HEAD	44	1
TO:HEAD+TO:LOC	43	1
REMOVE:HEAD+REMOVE:LOC	1	1

List of all constructions

AMPS

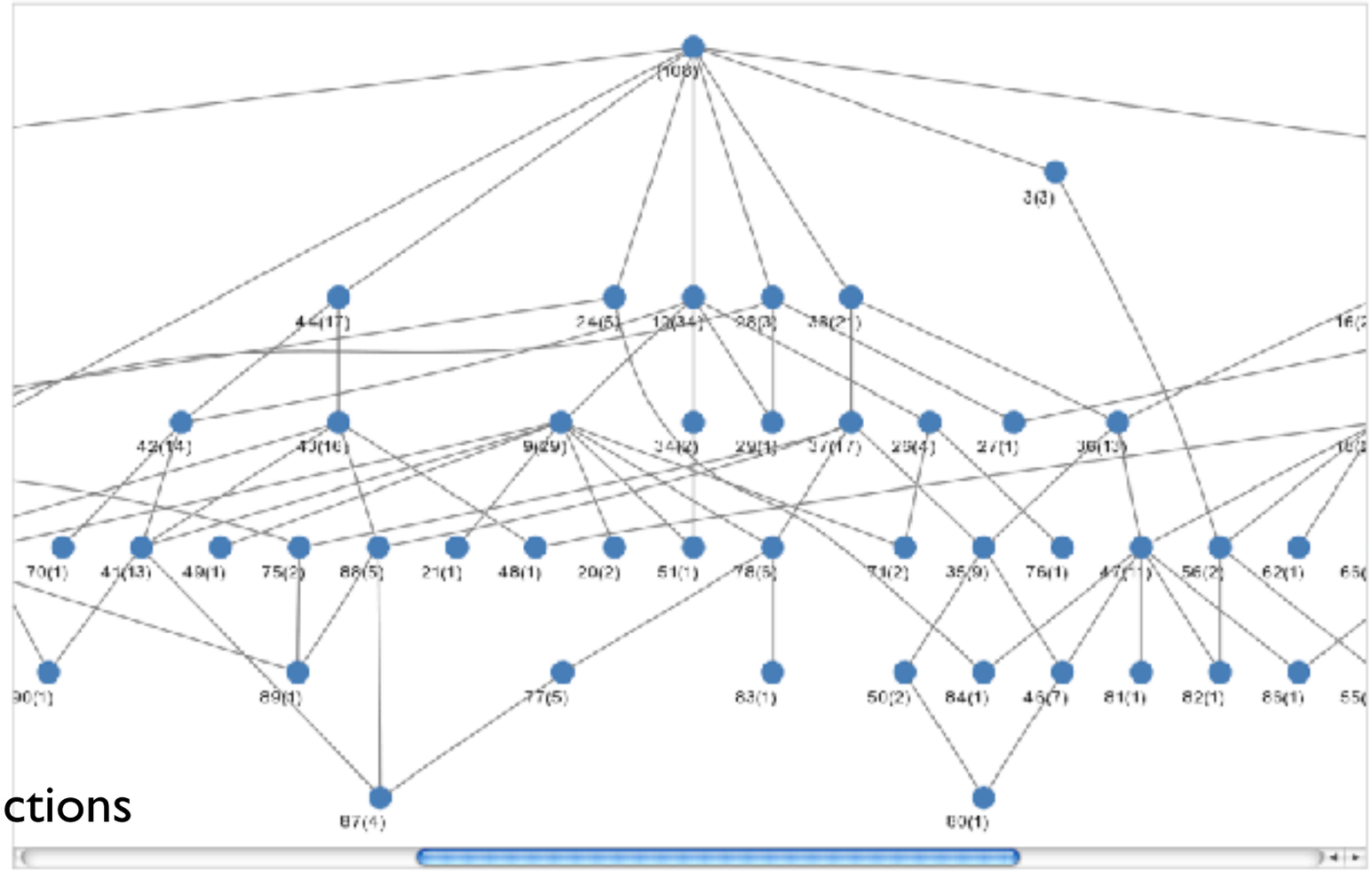
http://localhost:3000/s_construction/network?type=circle

Google

All Constructions

List of all construction instances

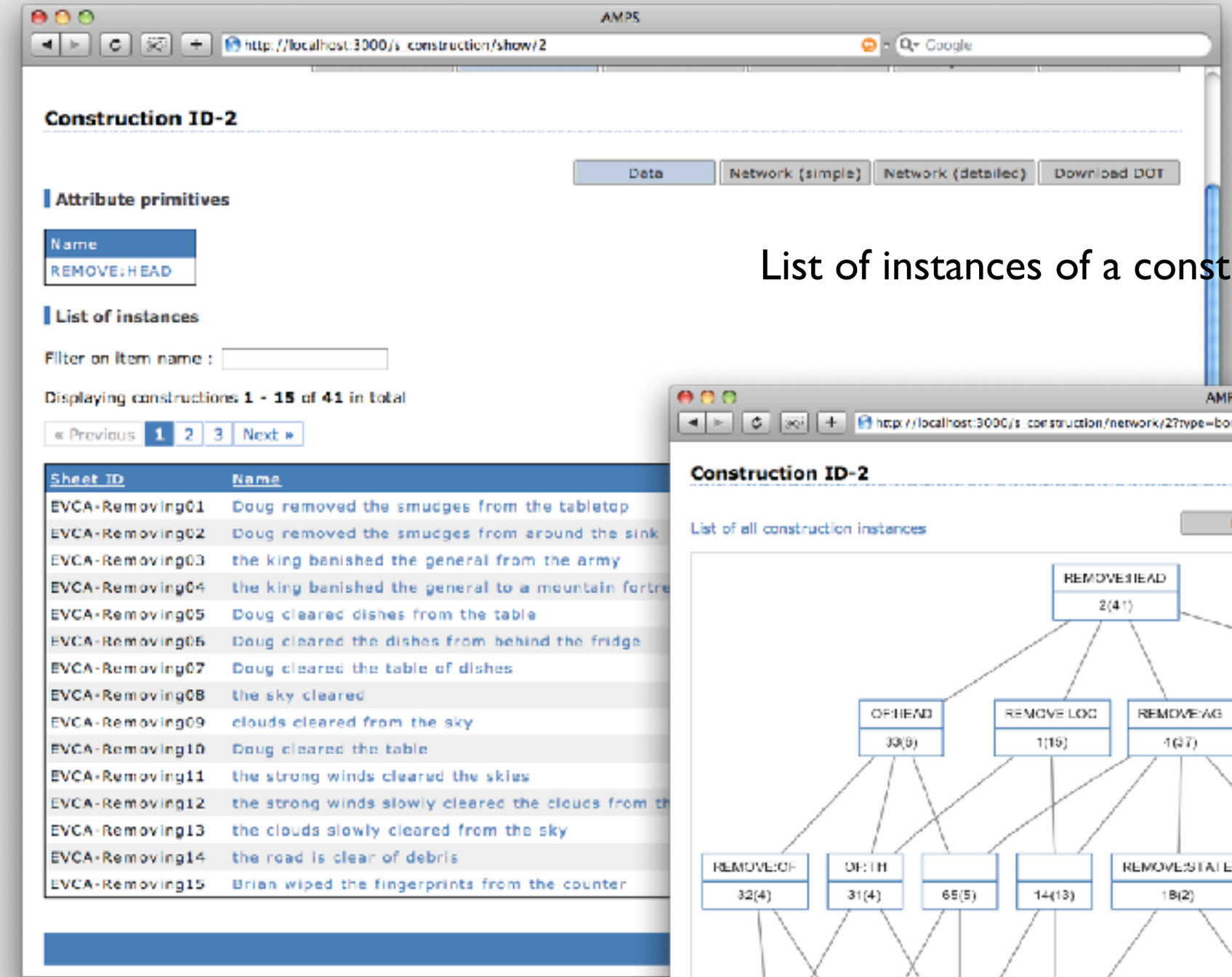
Data | Network (simple) | Network (detailed) | Download DOT

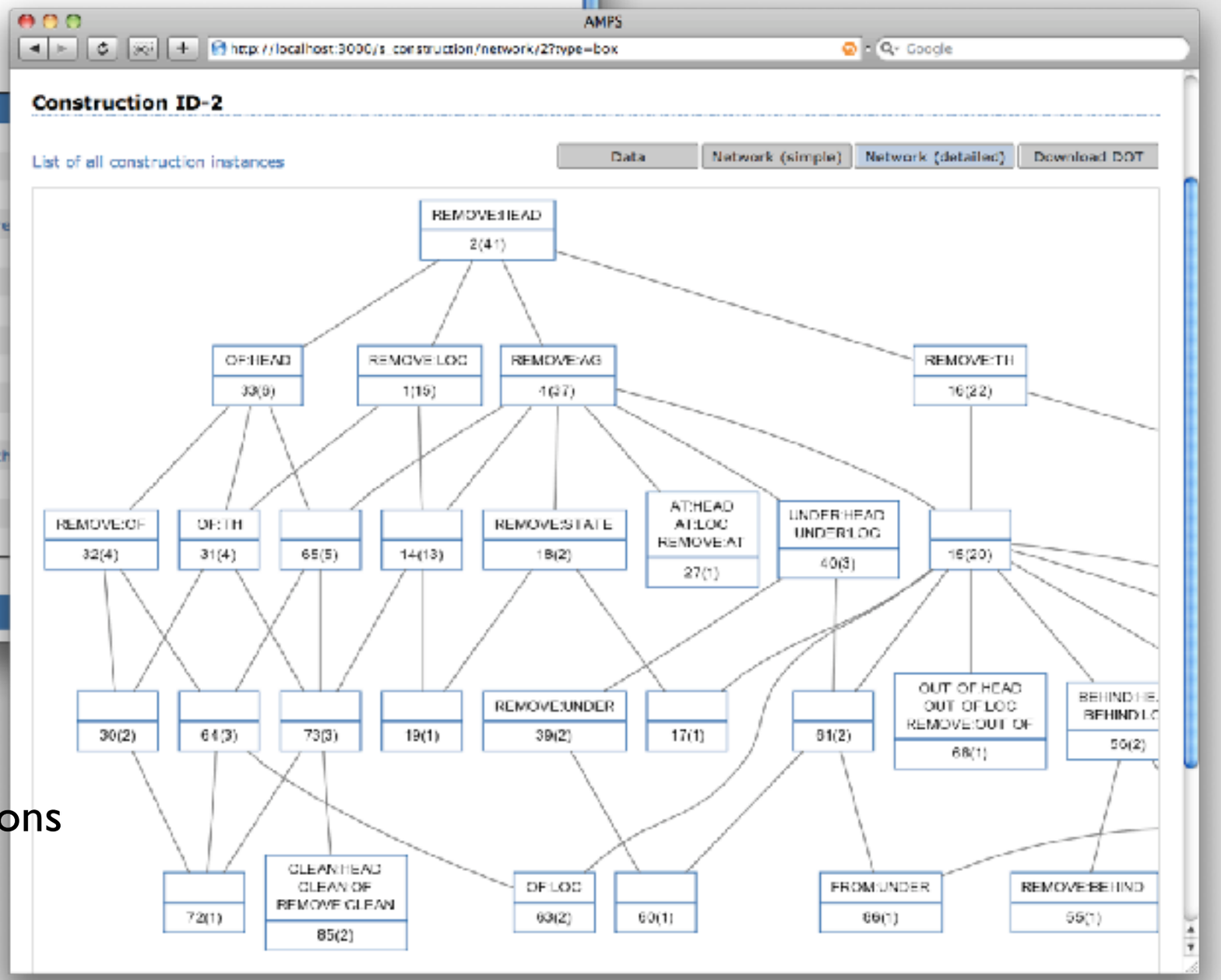


Network of all constructions

List of instances of a construction

Partial network of constructions





Background Issue I

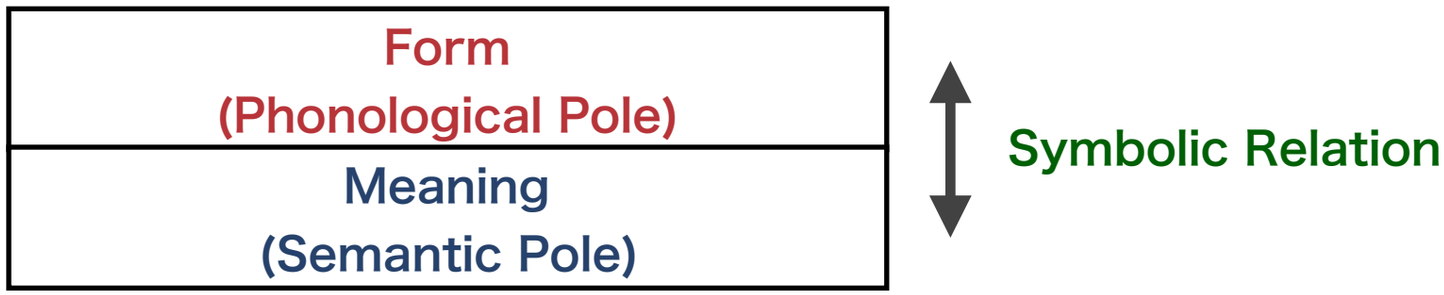
- Linguistics and OOD
  - Similarity between linguistic concepts and object-oriented software design (OOD)
  - Linguistic theories showing object-orientation

    - Goldberg's Construction Grammar
    - Filmore and Kay's Construction Grammar
    - Head Driven Phrase Structure Grammar (HPSG)
    - Langacker's Cognitive Grammar, etc.

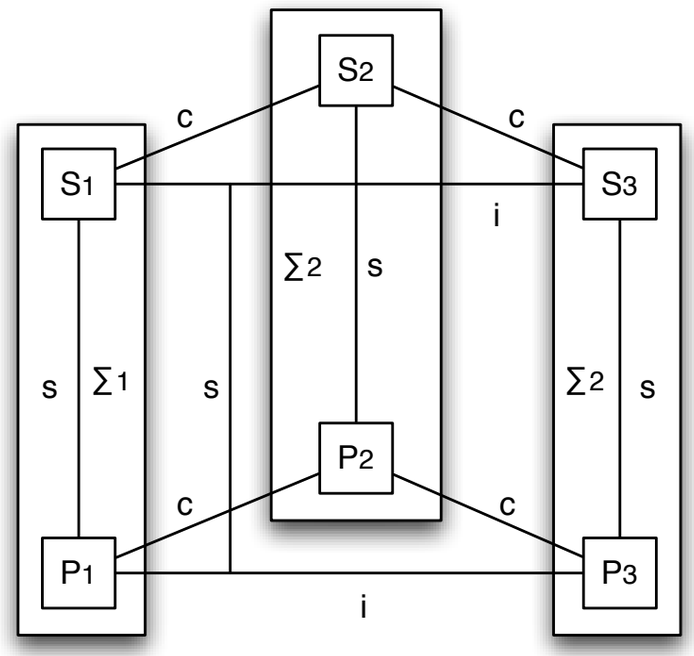
  - To create a system using an object-oriented programming language is easy, but what types of concepts are taken up as objects and how they are related to one another are more important

Background Issue 2

- Symbolic nature of linguistic structures
  - Saussurean tradition of linguistics, regarded as fundamental across disciplines and theories, especially so in non-derivational frameworks (CxG, HPSG, Cognitive Grammar).
  - The system needs to implement two separate domains, or poles.



c = composition
i = integration
s = symbolization



Langacker 2008: 162

Background Issue 3

- Creating a meta framework for corpus processing
  - The system accepts data sheets, each of which consists of a text sequence (typically a sentence) with various linguistic information annotated
  - As long as both design and practice of annotation are appropriate, there is no need to rely on statistical means to group items into grammatical categories ← this is exactly what linguists are always doing
  - The system offers a meta-framework on which various theories and analyses are tested for consistency and coverage.

Summary So Far

- Requirements for the system
  - Architecture
    an object-oriented architecture that reflects the symbolic thesis of language
  - Data Format
    the input data format has to be flexible enough to allow users to try out their own theories and analyses
  - Algorithm
    a method for detecting patterns in linguistic constructions from the object space while creating a hierarchical network for them

Solution 1 (Architecture)

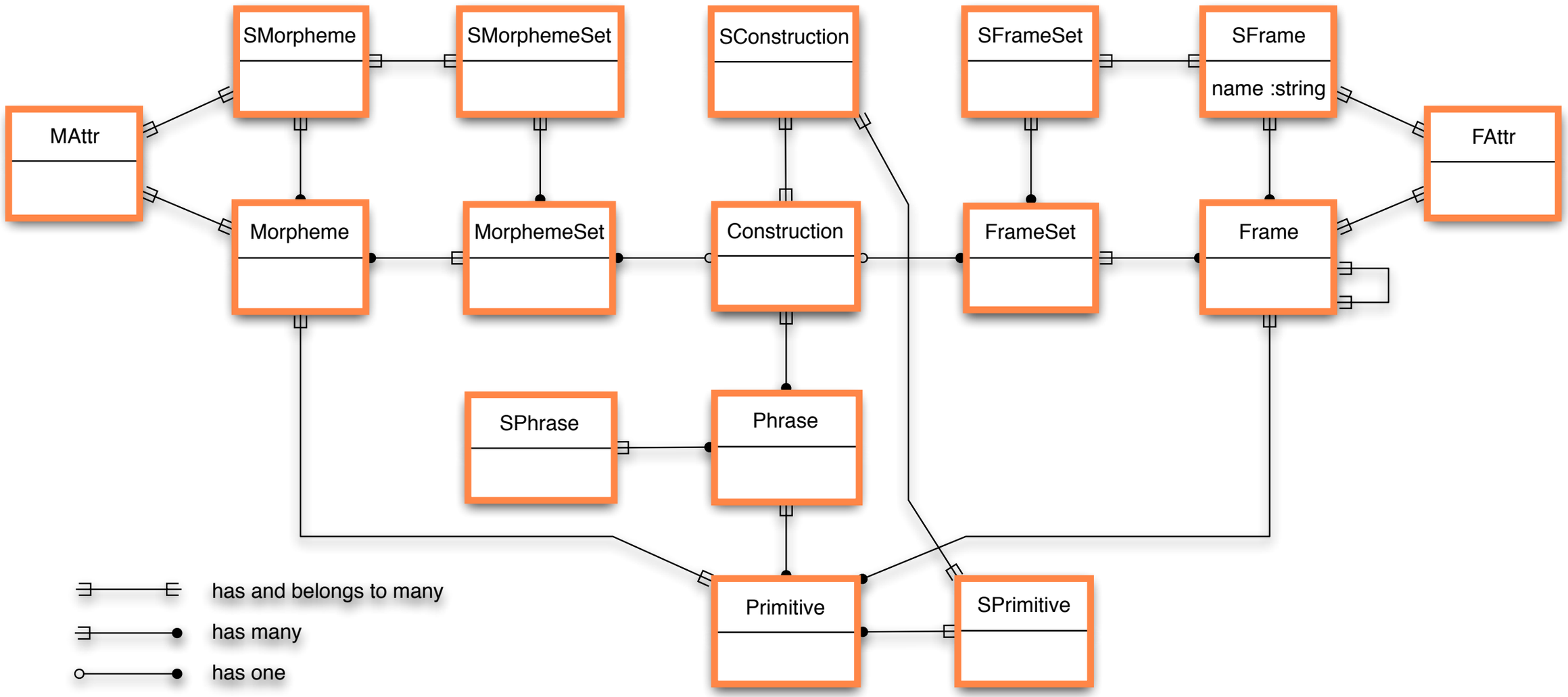
- Architecture of AMPS
  - Object-oriented design that reflects the symbolic structure of language
  - The object model of AMPS consists of 16 models
  - Most of them are categorized into either concrete category or schematic category

Schematic	SPrimitive	SMorpheme	SFrame	SPhrase	SMorphemeSet	SFrameSet	SConstruction
Concrete	Primitive	Morpheme	Frame	Phrase	MorphemeSet	FrameSet	Construction

  - A Construction-object represents a concrete instance of a construction (typically a sentence), while an SConstruction-object represents the schematic concept of a particular construction
  - Such a dual structure is also applied to the modeling of other objects

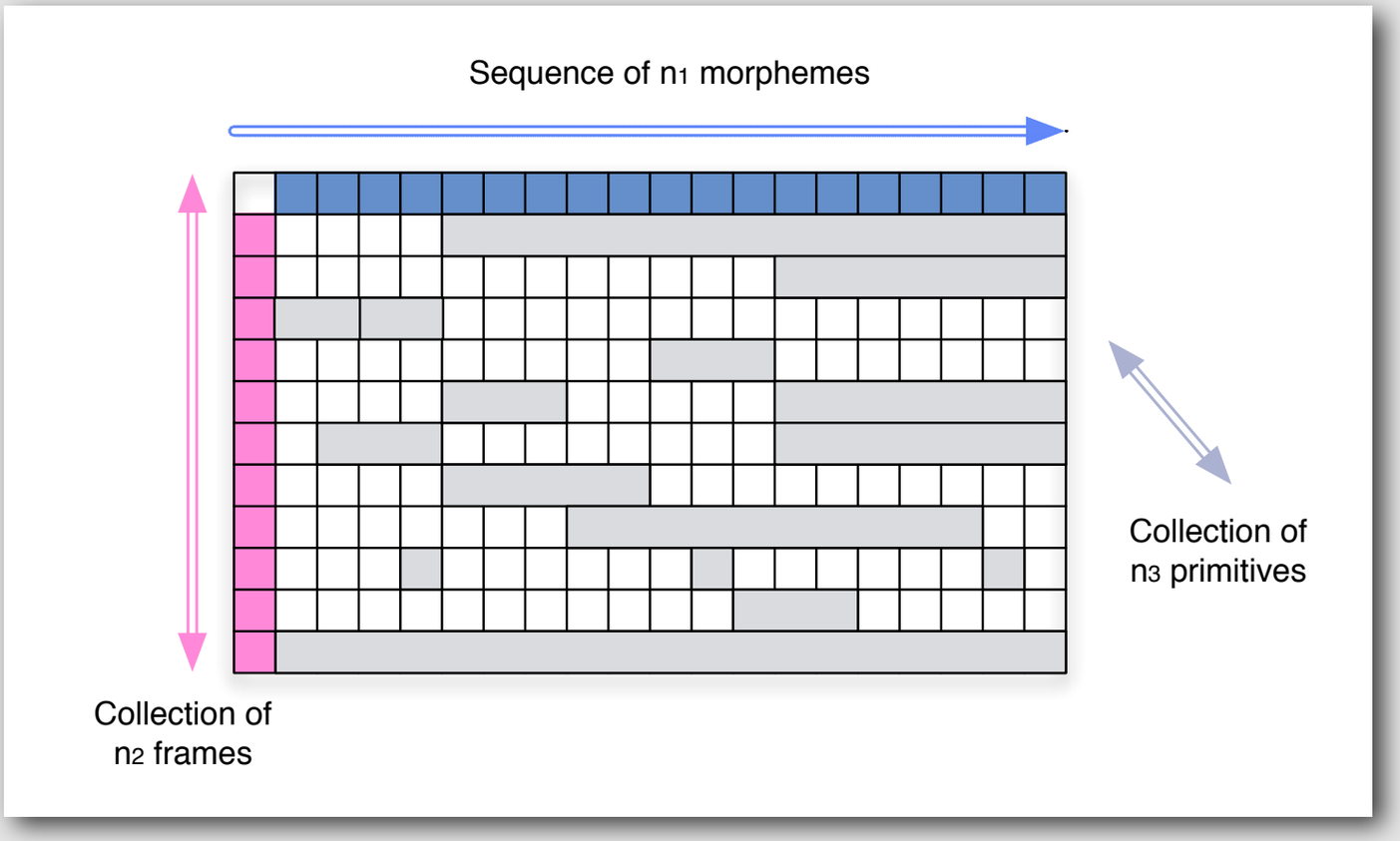
Architecture of AMPS

(generated semi-automatically from the real system)



Solution 2 (Data Format)

- Multi-layered matrix data format
  - Matrix of rows (frames) and columns (morphemes), with intersections having various kinds of values (primitives)
  - One data sheet corresponds to one construction instance (it could belong to more than one constructions)
  - Described in Microsoft Spreadsheet XML



EVCA-SendingAndCarrying24.xml

新規作成 開く 保存 プリント インポート コピー ペースト 書式 元に戻す やり直し オート SUM 昇順 降順 ギャラリー ツール 100% ズーム ヘルプ

シート グラフ SmartArt グラフィック ワードアート

	A	B	C	D	E	F	G	H	I	J	K	L
1	Sheet ID	EVCA-SendingAndCarrying24										
2												
3			M1	M2	M3	M4	M5	M6	M7	M8	M9	
4		Morphemes	Amanda	lugged	the	books	to	the	store	with	her	
5	MA1	original	Amanda	lug	the	book	to	the	store	with	her	
6	MA2	pos	N	V	D	N	P	D	N	P	N	
7	F1	SEND/CARRY	AG	HEAD	TH		TO			WITH		
8	F2	TO					HEAD	LOC				
9	F3	WITH								HEAD	LOC	
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												

AMPS XML Template +

標準表示 コマンド

EVCA-SendingAndCarrying24.xml

```
148    <Cell ss:StyleID="s21"><Data ss:Type="String">Sheet ID</Data></Cell>
149    <Cell ss:StyleID="s21"><Data ss:Type="String">EVCA-SendingAndCarrying24</Data></Cell>
150   </Row>
151   <Row ss:Index="3" ss:AutoFitHeight="0">
152    <Cell ss:StyleID="s22"><Data ss:Type="String"></Data></Cell>
153    <Cell ss:StyleID="s23"><Data ss:Type="String"></Data></Cell>
154    <Cell ss:StyleID="s22"><Data ss:Type="String">M1</Data></Cell>
155    <Cell ss:StyleID="s22"><Data ss:Type="String">M2</Data></Cell>
156    <Cell ss:StyleID="s22"><Data ss:Type="String">M3</Data></Cell>
157    <Cell ss:StyleID="s22"><Data ss:Type="String">M4</Data></Cell>
158    <Cell ss:StyleID="s22"><Data ss:Type="String">M5</Data></Cell>
159    <Cell ss:StyleID="s22"><Data ss:Type="String">M6</Data></Cell>
160    <Cell ss:StyleID="s22"><Data ss:Type="String">M7</Data></Cell>
161    <Cell ss:StyleID="s22"><Data ss:Type="String">M8</Data></Cell>
162    <Cell ss:StyleID="s22"><Data ss:Type="String">M9</Data></Cell>
163   </Row>
164   <Row ss:AutoFitHeight="0">
165    <Cell ss:StyleID="s22"><Data ss:Type="String"></Data></Cell>
166    <Cell ss:StyleID="s24"><Data ss:Type="String">Morphemes</Data></Cell>
167    <Cell ss:StyleID="s25"><Data ss:Type="String">Amanda</Data></Cell>
168    <Cell ss:StyleID="s25"><Data ss:Type="String">lugged</Data></Cell>
169    <Cell ss:StyleID="s25"><Data ss:Type="String">the</Data></Cell>
170    <Cell ss:StyleID="s25"><Data ss:Type="String">books</Data></Cell>
171    <Cell ss:StyleID="s25"><Data ss:Type="String">to</Data></Cell>
172    <Cell ss:StyleID="s25"><Data ss:Type="String">the</Data></Cell>
173    <Cell ss:StyleID="s25"><Data ss:Type="String">store</Data></Cell>
174    <Cell ss:StyleID="s25"><Data ss:Type="String">with</Data></Cell>
175    <Cell ss:StyleID="s25"><Data ss:Type="String">her</Data></Cell>
176   </Row>
177   <Row ss:AutoFitHeight="0">
178    <Cell ss:StyleID="s22"><Data ss:Type="String">MA1</Data></Cell>
179    <Cell ss:StyleID="s24"><Data ss:Type="String">original</Data></Cell>
180    <Cell ss:StyleID="s25"><Data ss:Type="String">Amanda</Data></Cell>
```

CAPS NUM

Create an AMPS Template

Source

```
General Settings
   num_of_morphs : 7

Morpheme Settings
   pos : NOUN, VERB, DET, PREP, ADJ, ADV, CONJ, COMP

Frame Settings
   PERFORMATIVE TYPE: STATEMENT, QUESTION, COMMAND
   RELATION TYPE: PROCESS, NON-PROCESSUAL-SIMPLIX, NON-PROCESSUAL-COMPLEX
```

Check Convert

The template source is **valid**

Preview

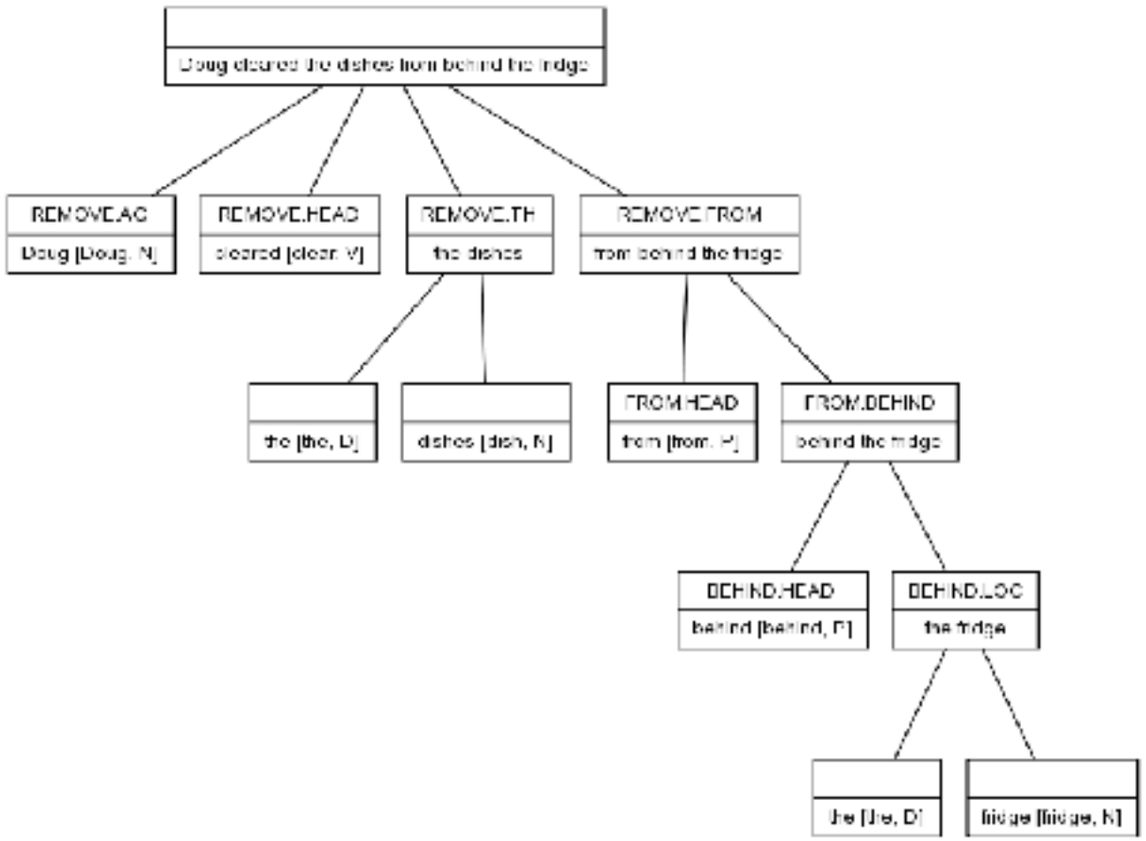
		M1	M2	M3	M4	M5	M6	M7	possible candidates
M	morphemes								
MA1	original								
MA2	pos								{NOUN, VERB, DET, PREP, ADJ, ADV, CONJ, COMP}
F1	PERFORMATIVE TYPE								{STATEMENT, QUESTION, COMMAND}
F2	RELATION TYPE								{PROCESS, NON-PROCESSUAL-SIMPLIX, NON-PROCESSUAL-COMPLEX}

- Annotation Example
  - A naive mixture of CxG, Cognitive Grammar and HPSG
  - Things (nominals) are considered participants of a frame with a semantic roles such as AG, TH, and LOC
  - Relations (VPs, PPs, etc) invoke frames as HEADs.
  - A frame may contain semantic roles and (subordinate) frames

(3) *Doug cleared the dishes from behind the fridge*

	A	B	C	D	E	F	G	H	I	J
1	Sheet ID	EVCA-Removing06								
2										
3			M1	M2	M3	M4	M5	M6	M7	M8
4		Morphemes	Doug	cleared	the	dishes	from	behind	the	fridge
5	MA1	original	Doug	clear	the	dish	from	behind	the	fridge
6	MA2	pos	N	V	D	N	P	P	D	N
7	F1	REMOVE	AG	HEAD	TH		FROM			
8	F2	FROM					HEAD	BEHIND		
9	F3	BEHIND						HEAD	LOC	

		M1	M2	M3	M4	M5	M6	M7	M8
	Morpheme	**Doug**	**cleared**	**the**	**dishes**	**from**	**behind**	**the**	**fridge**
	original	Doug	clear	the	dish	from	behind	the	fridge
	pos	N	V	D	N	P	P	D	N
F1	REMOVE	AG	HEAD	TH		FROM			
F2	FROM					HEAD	BEHIND		
F3	BEHIND						HEAD	LOC	



frames

Displaying **all 3** frames

Name	Count
REMOVE	1
FROM	1
BEHIND	1

primitives

Displaying **all 8** primitives

Name	Count
REMOVE:AG	1
REMOVE:HEAD	1
REMOVE:TH	1
REMOVE:FROM	1
FROM:HEAD	1
FROM:BEHIND	1
BEHIND:HEAD	1
BEHIND:LOC	1

morphemes

Displaying **all 7** morphemes

Name	Count
the	2
Doug	1
clear	1
dish	1
from	1
behind	1
fridge	1

phrases

Displaying **all 8** phrases

Name	Count
behind	1
the fridge	1
from behind the fridge	1
behind the fridge	1
Doug	1
the dishes	1
cleared	1
from	1

Solution 3 (Algorithm)

- Basic Concept
  - Constructions are defined as distinctive set of primitives, which are considered the minimal units in the symbolic relation of text.
    Cf. Radical Construction Grammar of Croft (2001)
  - Count only those that are not subsumed by other constructions
    Cf. Verb Island Hypothesis of Tomasello (2003)

1	*Fred*	*went*	*into*	*the*	*building*
2	*John*	*entered*	*into*	*the*	*store*
3	*Paul*	*fell*	*into*	*the*	*hole*

1	AG	GO	INTO	LOC
2	AG	GO	INTO	LOC
3	TH	GO	INTO	LOC

- # Formal Concept Analysis (FCA)
  - mathematically derives an ontology from a set of objects and attributes (Wille 2005)
  - the problem domain is represented as a formal context

    $\mathbb{K} = (G, M, I)$

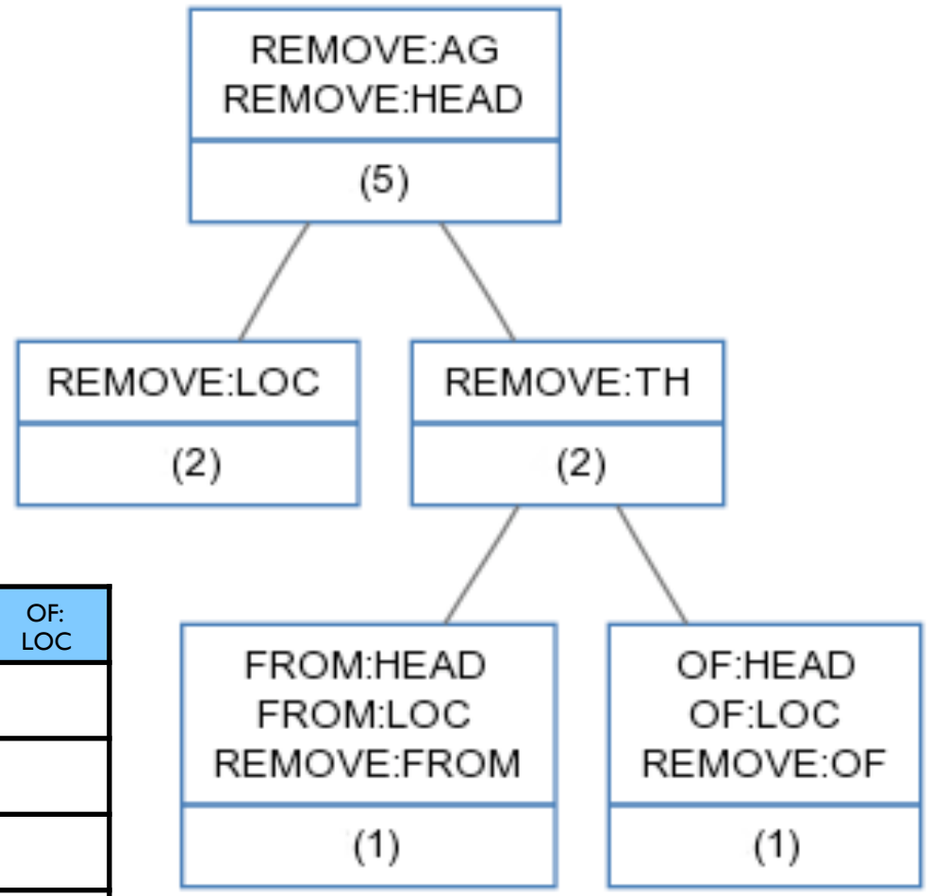
    G and M are sets; the elements of G are objects; the elements of M are attributes; and I is a binary relation between G and M.
  - a concept is defined as a pair of the extension and the intension: the extension consists of all objects belonging to the concept, while the intension comprises all attributes valid for all the objects in question
  - For two concepts (A, B) and (C, D), the subconcept-superconcept order is formalized as:

    $(A, B) \leq (C, D) :\Leftrightarrow A \subseteq C \ (\Leftrightarrow D \subseteq B)$

- How does FCA actually work on AMPS?
  - With the FCA method, constructions can be detected as concepts in the formal context with the input data as objects and primitives as attributes

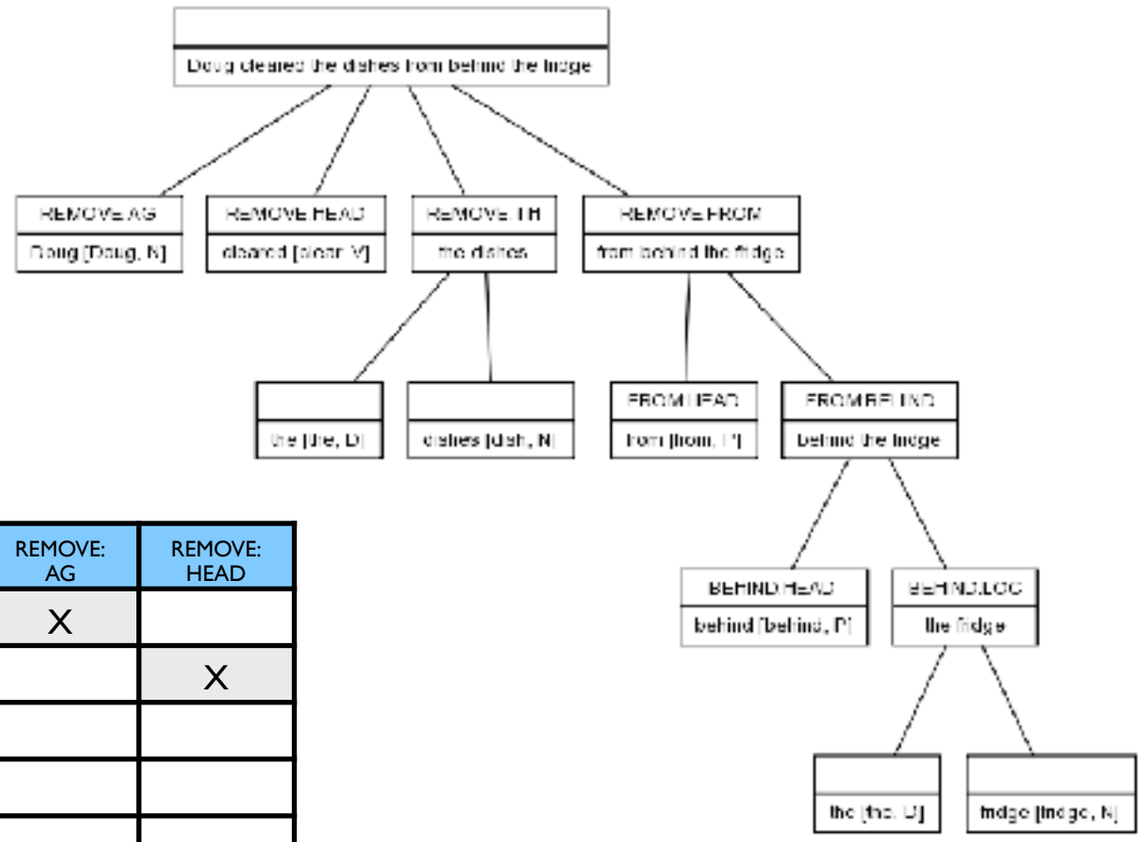
ID:1	Doug cleared the table
ID:2	Doug cleared dishes from the table
ID:3	Carla was shoveling
ID:4	Carla was shoveling the walk
ID:5	the doctor cured Pat of pneumonia

	REMOVE: AG	REMOVE: HEAD	REMOVE: TH	REMOVE: FROM	FROM: HEAD	FROM: LOC	REMOVE: LOC	REMOVE: OF	OF: HEAD	OF: LOC
ID:1	X	X								
ID:2	X	X					X			
ID:3	X	X	X	X	X	X				
ID:4	X	X	X							
ID:5	X	X	X					X	X	X



- Tree representation of a construction instance is also realized using FCA algorithm. (morphemes as objects, frames as attributes)

		M1	M2	M3	M4	M5	M6	M7	M8
	Morpheme	**Doug**	**cleared**	**the**	**dishes**	**from**	**behind**	**the**	**fridge**
	original pos	Doug N	clear V	the D	dish N	from P	behind P	the D	fridge N
F1	REMOVE	AG	HEAD	TH		FROM			
F2	FROM					HEAD	BEHIND		
F3	BEHIND						HEAD	LOC	



	BEHIND: LOC	REMOVE: TH	FROM: HEAD	FROM: BEHIND	REMOVE: FROM	FROM: BEHIND	REMOVE: AG	REMOVE: HEAD
Doug							X	
cleared								X
the		X						
dishes		X						
from			X			X		
behind				X	X	X		
the	X				X	X		
fridge	X				X	X		

Conclusion

- What is AMPS?
  - A meta-framework for processing corpus data
  - Feature of AMPS
    - Has an object-oriented and symbol-oriented architecture
    - Accept data in multi-layered matrix format
    - Converts data into networks of linguistic constructions using FCA algorithms
- Applications to linguistic research
  - A testing platform for a new new theory/analysis
  - Statistical data analysis
  - Building a frame database
- Future plan
  - Further development, testing, optimization, and documentation
  - Public release → as an open source project, early next year (?)

Acknowledgement

The author is grateful to Kow Kuroda at NICT for the basic idea of developing a multi-layered data processing system, and for valuable comments and suggestions that considerably improved the present system.

Comments welcome

Email: yohasebe@gmail.com

References

Baker, Collin, Charles J. Fillmore, and Beau Cronin. 2003. The structure of the FrameNet database. *International Journal of Lexicography* 16: 281–96.

Bergen, Benjamin K., and Nancy Chang. 2005. Embodied construction grammar in simulation-based language understanding. In J.-O. Östman and M. Fried (eds.), *Construction Grammars: Cognitive Grounding and Theoretical Extensions*, 147-90. Amsterdam: John Benjamins.

Boas, Hans C. 2005. Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography* 18, 445-78.

Cook, Walter A. 1998. *Case Grammar Applied*. Dallas, TX: Summer Institute of Linguistics.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

De Beule, Joachim. and Luc Steels. 2005. Hierarchy in fluid construction grammar. In U. Furbach (ed.), *Proceedings of the 28th Annual German Conference on AI*, Lecture Notes in Artificial Intelligence (Vol. 3698), 1-15. Berlin; Heidelberg.

Fillmore, Charles. J., Christopher R. Johnson, and Miriam R. L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography* 16: 235–51.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Kay, Paul. 2002 An informal sketch of a formal architecture for construction grammar. *Grammars* 5: 1-19.

Kay, Paul, and Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: the *what's x doing y?* construction. *Language* 75: 1-33.

Kuroda, Kow, and Hitoshi Isahara. 2005. Proposing the multilayered semantic frame analysis of text. *Proceedings of the 3rd International Workshop of Generative Approaches to the Lexicon*, 124-133.

Kuroda, Kow, Masao Utiyama, and Hitoshi Isahara. 2006. Getting deeper semantics than Berkeley FrameNet with MSFA. *Proceedings of the Language Resources and Evaluation 2006*, P26-EW.

Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar*, Vol. 2. Stanford, CA: Stanford University Press.

Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation.* Chicago: University of Chicago Press.

Ohara, Kyoko H., Seiko Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito, and Shun Ishizaki. 2006. Frame-based contrastive lexical semantics and Japanese FrameNet: The case of RISK and *kakeru.* Paper presented at the Fourth International Conference on Construction Grammar, Tokyo, Japan.

Östman, Jan-Ola and Mirjam Fried. 2005. The cognitive grounding of construction grammar. In J.-O. Östman and Mirjam Fried (eds.), *Construction Grammars: Cognitive Grounding and Theoretical Extensions*, 1-16. Amsterdam: John Benjamins.

Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Sag, Ivan A., Thomas Wasow, and Emily M. Bendler. *Syntactic Theory: A Formal Introduction*, Second Edition. Stanford, CA: CSLI Publications.

Steels, Luc, and Joachim De Beule. 2006. A (very) brief introduction to fluid construction grammar. *Proceedings of the 3rd International Workshop on Scalable Natural Language*. Document available online at http://acl.ldc.upenn.edu/W/W06/W06-3510.pdf

Tomasello, Michael. 2003. *Creating a Language: A Usage-Based Theory of Language Acquisition.* Cambridge, MA: Harvard University Press.

Wille, Rudolf. 2005. Formal concept analysis as mathematical theory of concepts and concept hierarchies. In B. Ganter, G. Stumme, and R. Wille (eds.), *Formal Concept Analysis: Foundations and Applications*, 1-33. Berlin: Springer.

Wolff, Karl E. 1994. A first course in formal concept analysis: How to understand line diagrams. In F. Faulbaum (ed.), *SoftStat '93*, Advances in Statistical Software 4, 429-38. Stuttgart: Gustav Fischer Verlag.