

# No Sooner Said Than Done? Testing Incrementality of Semantic Interpretations of Spontaneous Speech

Michaela Atterer, Timo Baumann, David Schlangen

Department for Linguistics, University of Potsdam, Germany

## I. Introduction

Background: In highly interactive settings, **Incremental Spoken Dialogue Systems** are preferred over non-incremental systems as they **react faster and more naturally** (Aist et al. 2007).

- need to build up syntactic and semantic structure on the fly while the user is still speaking.
- needs theory-neutral assessment of the quality of incrementally built semantic structure.

## Research Goals

- present generic measures to evaluate incremental semantics construction
- focus is on measuring the incrementality (not necessarily the quality of the non-incremental result)
- show and analyze the performance of our specific module on a specific corpus

## II. Incremental Semantics Construction

- Approaches vary by **strictness of incrementality** (Nivre, 2004) vs. ambiguity of structures
- classical **trade-off**: slow and precise or quick and vague

## Previous Work

Aist et al., (2006) and Bücher et al. (2002) generate (partial) hypotheses once semantics can be constrained to a small set:

“move a large triangle to ...”  
↓  
move(X, Y)      move(triangle, Y)

Schuler (2002), Brick and Scheutz (2007), and others generate (all) structures that might possibly match in the future:

“move a large triangle to ...”  
↓  
move(triangle ∨ square ∨ circle ∨ ...)

Neither of the previous work evaluates the incrementality of the semantic interpretation on a corpus.

Partly, evaluation is intrinsic to the used semantics construction mechanisms.

## III. Evaluating Incremental Semantics Construction

There is **no incremental gold-standard!**

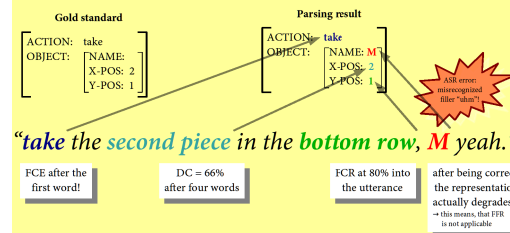
- in non-incremental evaluation, we can just compare to the human annotated gold-standard
- there is no (cannot be?) annotation of **what should be known up to a certain word in the utterance**
- All in all, we want the final result (or parts thereof) as soon as possible.

## Measures

The measures we propose relate **relative position** in the utterance and **comparison to the non-incremental gold**.

Also, we differentiate **both partial and complete success**:

- **first correct element (FCE)**:  
When is the first element of the representation correct?
- **first correctly-filled representation (FCR)**:  
When (in percent) do we first match the gold-standard?
- **first finally correctly-filled representation (FFR)**:  
When does the correct representation not change anymore?
- **degree of correctness (DC)** at a certain time:  
The percentage of elements correctly filled on average.

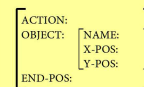


## IV. Application

### Domain

Non-interactive explanations of puzzle-piece placements in the Pentomino domain. Data from (Siebert and Schlangen, 2008).

- manually transcribed with a 5-slot frame-semantics
- 500 utterances: 100 for grammar development, 400 in the evaluation
- 332 test utterances (68 utterances had completely empty frames)
- we distinguish 171 short utterances (≤ 10 words) and 161 long utterances (> 10 words)

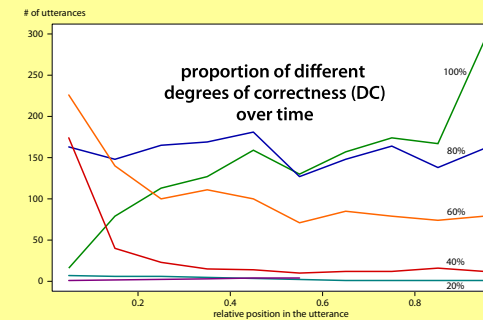
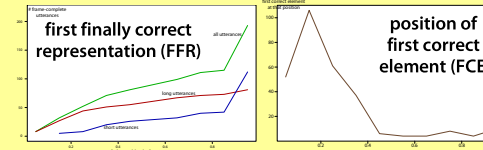


## Our Incremental Semantics Component

We use RUBISC, the Robust Unification-Based Incremental Semantic Chunker (Atterer and Schlangen, 2009).

- based on the idea of semantic units (Selkirk, 1984), which correspond to phonological phrases
- collect word material until there is enough information to change the state of the semantic frame
- contentful units are defined in a grammar via regular expressions
- words are consumed until a unit is complete and the corresponding slots in the frame are filled
- selectional restrictions can be modelled by filling slots with special “blocker” values

## Results



- interpretation can often be completed before the utterance is over  
→ this is especially true for long utterances.
- parts of the interpretation are already first correct (FCE) in the beginning of an utterance  
→ this could e.g. be used to prepare possible system responses.
- high degrees of correctness increase with time, low decrease
- considerable knowledge after only 40% of the utterance
- relative stability in DC between 40% to 80% of the utterance  
→ the first few and the final word in the utterance are most important.

## V. Discussion

We have defined measures to evaluate the incrementality of semantic components against a non-incremental gold-standard:

- measures seem to capture meaningful aspects of incremental semantic interpr.,
- measures are generic enough and allow to compare components with differing approaches to semantic interpretation.

We have evaluated our semantic component and found that incremental semantic interpretation is worthwhile:

- considerable knowledge with only parts of the utterance available.

Our results are obviously limited to corpus and semantics:

- standardized corpora and annotations are needed to compare different approaches.

## Bibliography

- [1] G. Aist, J. Allen, E. Campana, G. Gallo, S. Stoness, M. Swift, and M. Tanenhaus, “Incremental understanding in human-computer dialogue and experimental evidence for advantages over nonincremental methods”, in *Dialog 2007*, Trento, Italy, 2007.
- [2] J. Nivre, “Incrementality in deterministic dependency parsing”, in *Proc. of the ACL Workshop Incremental Parsing: Bringing Engineering and Cognition Together*, Barcelona, Spain, 2004.
- [3] G. Aist, S. Stoness, and J. Allen, “Steps towards incremental semantics for spoken dialog systems”, in *MCLC-2006*, Urbana, USA, 2006.
- [4] K. Bücher, G. Görz, and B. Ludwig, “Coreg tabs: Incremental semantic composition”, in *Proc. of KI-2002 Workshop on Applications of Description Logics*, Aachen, Germany, 2002.
- [5] H. Kamp and U. Reyle, *From Discourse to Logic*. Dordrecht: Kluwer, 1993.
- [6] W. Schuler, “Interleaved semantic interpretations in environment-based parsing”, in *Proc. of COLING-2002*, Taipei, Taiwan, 2002.
- [7] T. Brick and M. Scheutz, “Incremental natural language processing for HRI”, in *Proc. of HRI 07*, New York, USA, 2007.
- [8] C. Rosé, A. Roque, and D. Bhembe, “An efficient incremental architecture for robust interpretation”, in *Proc. of HLT 2007*, San Francisco, USA, 2002.
- [9] S. Stoness, J. Treadall, and J. Allen, “Incremental parsing with reference intraction”, in *ACL Workshop on Incremental Parsing*, Barcelona, Spain, 2004.
- [10] M. Atterer and D. Schlangen, “Rubisc – a robust unification-based incremental semantic chunker”, in *Proc. of SRSI-2009*, Athens, Greece, 2009.
- [11] E. Selkirk, *Phonology and Syntax: The relation between sound and structure*. Cambridge, USA: MIT Press, 1984.
- [12] M. Nespore and L. Vogel, *Prosodic Phonology*. Dordrecht: Foris Publications, 1986.
- [13] A. Siebert and D. Schlangen, “A simple method for resolution of definite reference in a shared visual context”, in *Proc. of SIGDial*, Columbus, Ohio, 2008.

## Further Information

Please contact us via e-mail at  
{[atterer](mailto:atterer@ling.uni-potsdam.de) | [timo](mailto:timo@ling.uni-potsdam.de) | [das](mailto:das@ling.uni-potsdam.de)}@ling.uni-potsdam.de



This work is funded by a DFG grant in the Emmy Noether programme.