Sapienza Università di Roma

UiO : Universitetet i Oslo

Non-Sentential Utterances in Dialogue: Experiments in classification and interpretation

Paolo Dragone
Sapienza University of Rome

Pierre Lison
University of Oslo

Non-Sentential Utterances

What is an NSU?

Non-sentential utterance (NSU): utterance without a complete sentential form that conveys a full clausal meaning given the dialogue context.

A: How do you actually fell about that?
B: **Not too happy.**

A: They wouldn't do it, no.
B: **Why?**

A: [...] then across from there to there.
B: **From side to side.**

NSU taxonomy/corpus

NSU class	Total	%
Plain Acknowledgment	599	46.1
Short Answer	188	14.5
Affirmative Answer	105	8.0
Repeated Acknowledgment	86	6.6
Clarification Ellipsis	82	6.3
Rejection	49	3.7
Factual Modifier	27	2.0
Repeated Affirmative Answer	26	2.0
Helpful Rejection	24	1.8
Check Question	22	1.7
Sluice	21	1.6
Filler	18	1.4
Bare Modifier Phrase	15	1.1
Propositional Modifier	11	0.8
Conjunct	10	0.7
Total	1283	100.0

Motivation

- About 10% of the total number of utterances in dialogues are NSUs.
- NSUs are highly context-dependent.
- Variety of NSU classes with distinct linguistic structures and pragmatic functions.
- The resolution of NSUs requires deep semantic understanding.
- The interpretation of NSUs is still an understudied problem.

An Active Learning approach to the classification of NSUs

Approach

- Baseline, replicated approach from Fernández et al. (2007).
- Extended feature-set, 23 new features.
- Active learning techinques used to cope with scarsity of labeled data and strong class imbalance.
- Additional unlabeled data extracted from the BNC via simple heuristics.
- 100 newly labeled instances added to the training-set.

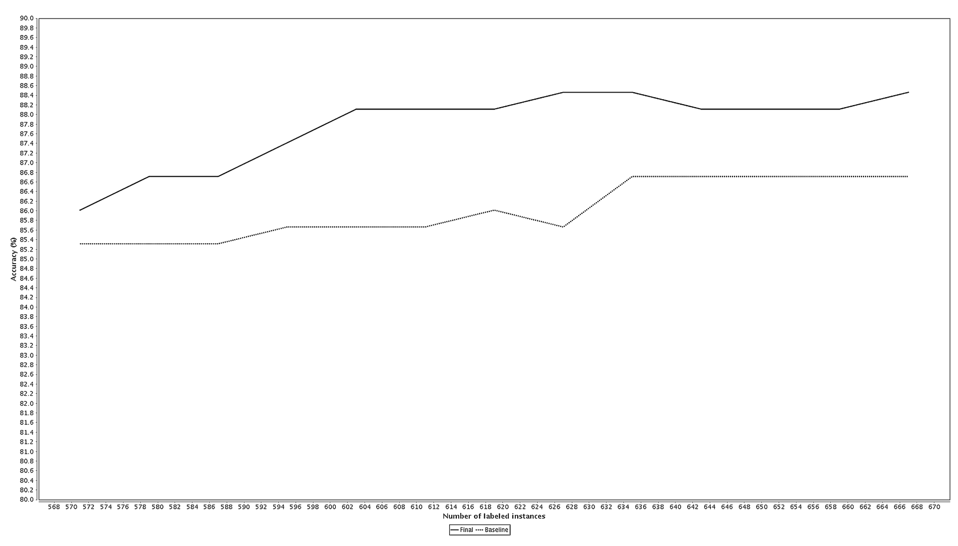
Baseline feature-set

- NSU features: syntactic and lexical properties of the NSU.
- Antecedent features: syntactic and lexical properties of the antecedent.
- Similarity features: similarity measures between NSU and antecedent.

Extended feature-set

- POS-level features: shallow syntactic properties using POS-tags.
- Phrase-level features: phrasal syntactic patterns.
- Dependency features: dependency syntactic patterns.
- Turn taking features: patterns in the conversational flow.
- Similarity features: longest common subsequences of words and POS tags.

Learning curve



Accuracy of the classifier throughout the active learning process using the baseline feature-set (dashed line) and the final feature-set (solid line).

Empirical results

Training set (feature set)	Accuracy	Precision	Recall	F_1-score
Train-set (baseline feature set)	0.881	0.884	0.881	0.875
Train-set (extended feature set)	0.899	0.904	0.899	0.896
Train-set + AL (baseline feature set)	0.883	0.893	0.883	0.880
Train-set + AL (extended feature set)	**0.907**	**0.913**	**0.907**	**0.905**

Interpretation of NSUs using probabilistic rules

Approach

- Dialogue context modelling inspired by Ginzburg (2012).
- NSU resolution procedures based on Fernández (2006).
- Context update rules reinterpreted as *probabilistic rules* (Lison, 2015).
- Rules implemented with **OpenDial**.
- Ongoing evaluation of the rules over transcripts from the Communicator dataset.
- Proof of concept for a framework for the interpretation of NSUs based on the probabilistic rules formalism.

Why use probabilistic rules?

Probabilistic update of the dialogue state

A: Is Jack coming to the party next Saturday?
B: **Probably. / Unlikely.**

$$\begin{array}{l}
\forall q, \mathbf{y} \\
\textbf{if } (\texttt{nsu}_{\texttt{u}} = \mathit{PropMod} \wedge \texttt{MaxQUD.q} = q(\mathbf{y})) \textbf{ then} \\
\quad \left\{ P \begin{pmatrix} \texttt{a}_{\texttt{u}} \leftarrow \mathit{Assert}(\mathit{PropRel}_{\texttt{u}_{\texttt{u}}}(q)(\mathbf{y})) \\ \texttt{new-fec} \leftarrow \texttt{MaxQUD.fec} \end{pmatrix} = 1 \right.
\end{array}$$

As in Fernández (2006), propositional modifiers are resolved with a lexical relation *PropRel*.

The update of Facts following the acceptance of the above utterance can be handled probabilisically.

$$\begin{array}{l}
\forall p, \mathbf{y} \\
\textbf{if } (\texttt{a}_{\texttt{m}} = \mathit{Accept}(\mathit{probably}(p)(\mathbf{y}))) \textbf{ then} \\
\quad \left\{ P(\texttt{facts} \leftarrow \texttt{facts} \cup \{p(\mathbf{y})\}) = 0.75 \right. \\
\textbf{else if } (\texttt{a}_{\texttt{m}} = \mathit{Accept}(\mathit{unlikely}(p)(\mathbf{y}))) \textbf{ then} \\
\quad \left\{ P(\texttt{facts} \leftarrow \texttt{facts} \cup \{p(\mathbf{y})\}) = 0.25 \right.
\end{array}$$

Although the probabilities are handcrafted here, they could be learned from actual data.

Probabilistic account of ambiguities

A: One of our salesmen acquired a new client.
B: **Who?** (= the salesman?/the client?)

$$\begin{aligned}
\texttt{MaxQUD.q} &= \mathit{acquired}(X_1, X_2) \\
\texttt{MaxQUD.fec} &= \{\mathit{salesman}(X_1), \mathit{client}(X_2)\}
\end{aligned}$$

$$\begin{array}{l}
\forall q, x, \mathbf{y}, p_i, \mathbf{y}_i \\
\textbf{if } (\texttt{nsu}_{\texttt{u}} = \mathit{Sluice} \wedge \text{“who”} \in \texttt{u}_{\texttt{u}} \wedge \texttt{MaxQUD.q} = q(x, \mathbf{y}) \wedge \\
\quad \{p_1(x, \mathbf{y}_1), \ldots, p_n(x, \mathbf{y}_n)\} \subseteq \texttt{MaxQUD.fec}) \textbf{ then} \\
\quad \left\{ P \begin{pmatrix} \texttt{a}_{\texttt{u}} \leftarrow \mathit{Ask}(\mathit{named}(x, \hat{x})) \\ \texttt{new-fec} \leftarrow \{p_1(x, \mathbf{y}_1), \ldots, p_n(x, \mathbf{y}_n)\} \cup \{\mathit{person}(x)\} \end{pmatrix} = 1 \right.
\end{array}$$

A *who*-sluice is resolved by asking the identity of the person underspecified in the antecedent.

If there is more than one focal constituent the resolution is ambiguous but it is automatically handled probabilistically.

$$\begin{aligned}
\texttt{a}_{\texttt{u}} &= \begin{cases} \mathit{Ask}(\mathit{named}(X_1, X_3)) & \text{with probability } p = 0.5 \\ \mathit{Ask}(\mathit{named}(X_2, X_3)) & \text{with probability } p = 0.5 \end{cases} \\
\texttt{new-fec} &= \begin{cases} \{\mathit{salesman}(X_1), \mathit{person}(X_1)\} & \text{with probability } p = 0.5 \\ \{\mathit{client}(X_2), \mathit{person}(X_2)\} & \text{with probability } p = 0.5 \end{cases}
\end{aligned}$$

A more sophisticated representation could use some notion of *saliency* for each constituent.