The Empirical Grammar of Market Conversations

A Neuro-Symbolic Reconstruction

Paul Koop

1994–2026

Abstract

This paper reconstructs the empirical grammar of eight sales conversations recorded at Aachen market square in 1994, using the Algorithmic Recursive Sequence Analysis (ARS) framework. The analysis proceeds from raw transcripts to terminal symbol chains, transition counting, probability induction, and finally to a probabilistic context-free grammar (PCFG) with empirically optimized transition probabilities. I provide an interpretive analysis of the learned probabilities, distinguishing between *constitutive rules* (structural constraints that define well-formedness) and *statistical regularities* (empirical frequencies that reflect contingent patterns). The grammar reveals both ritualized sequences (greetings, farewells) and strategic options (upselling loops, restarts). The paper concludes with a discussion of how the grammar can be implemented in neuro-symbolic frameworks, bridging qualitative interpretation and computational execution.

Contents

1 Introduction: From Raw Data to Formal Grammar

The Algorithmic Recursive Sequence Analysis (ARS) rests on a simple yet powerful idea: social interactions leave physical traces (audio recordings) that can be transcribed, interpreted, and transformed into formal grammars. This paper documents the complete pipeline from raw transcripts to an empirically optimized grammar, using eight sales conversations recorded at Aachen market square in June/July 1994.

The contribution is threefold:

1. **Empirical**: I present the full optimized grammar induced from the eight transcripts, with all transition probabilities.
2. **Interpretive**: I analyze the learned probabilities, distinguishing between constitutive rules (structural necessities) and statistical regularities (empirical contingencies).
3. **Methodological**: I discuss how the grammar can be implemented in neuro-symbolic frameworks, bridging qualitative interpretation and computational execution.

2 Data and Interpretation

2.1 The Eight Transcripts

The empirical material comprises eight transcripts of sales conversations recorded at Aachen market square in June/July 1994. The transcripts vary in length, completeness, and conversation type:

Table 1: Corpus of market conversations

Text	Date	Location	Participants
T1	28.06.1994	Butcher shop	Seller (f), Customer
T2	28.06.1994	Cherry stall	Seller (m), C1, C2
T3	28.06.1994	Fish stall	Seller (m), Customer
T4	28.06.1994	Vegetable stall	Seller (m), Customer
T5	26.06.1994	Vegetable stall	Seller (m), C1, C2
T6	28.06.1994	Cheese stall	Seller (m), C1
T7	28.06.1994	Candy stall	Seller (m), Customer
T8	09.07.1994	Bakery	Seller (f), Customer

2.2 Terminal Symbol Assignment

Each utterance was assigned a terminal symbol from a predefined category system, developed through sequential micro-analysis following objective hermeneutics (Oevermann et al., 1979). The twelve terminal symbols are:

Table 2: Terminal symbols and their meanings

Symbol	Meaning
KBG	Customer greeting
VBG	Seller greeting
KBBd	Customer need (concrete)
VBBd	Seller inquiry
KBA	Customer response
VBA	Seller reaction
KAE	Customer inquiry
VAE	Seller information
KAA	Customer completion
VAA	Seller completion
KAV	Customer farewell
VAV	Seller farewell

The resulting terminal symbol chains for the eight transcripts are:

```
T1: KBG, VBG, KBBd, VBBd, KBA, VBA, KBBd, VBBd, KBA, VAA, KAA, VAV, KAV
T2: VBG, KBBd, VBBd, VAA, KAA, VBG, KBBd, VAA, KAA
```

```
T3: KBBd, VBBd, VAA, KAA
T4: KBBd, VBBd, KBA, VBA, KBBd, VBA, KAE, VAE, KAA, VAV, KAV
T5: KAV, KBBd, VBBd, KBBd, VAA, KAV
T6: KBG, VBG, KBBd, VBBd, KAA
T7: KBBd, VBBd, KBA, VAA, KAA
T8: KBG, VBBd, KBBd, VBA, VAA, KAA, VAV, KAV
```

3 The Optimized Grammar

3.1 Induction Method

The grammar was induced by counting transitions across all eight transcripts and normalizing to probabilities:

$$P(\sigma_j \mid \sigma_i) = \frac{\text{count}(\sigma_i \to \sigma_j)}{\sum_k \text{count}(\sigma_i \to \sigma_k)}$$

The resulting probabilities were iteratively refined by generating artificial chains and comparing their frequency distributions to the empirical data. The final grammar achieved a correlation of $r = 0.925$ between empirical and generated frequencies.

3.2 The Optimized Probabilistic Grammar

Table 3: Optimized transition probabilities

Start symbol	Following symbols with probabilities
KBG	VBG (0.667), VBBd (0.333)
VBG	KBBd (1.0)
KBBd	VBBd (0.667), VAA (0.167), VBA (0.167)
VBBd	KBA (0.444), VAA (0.222), KBBd (0.222), KAA (0.111)
KBA	VBA (0.5), VAA (0.5)
VBA	KBBd (0.5), KAE (0.25), VAA (0.25)
VAA	KAA (0.857), KAV (0.143)
KAA	VAV (0.75), VBG (0.25)
VAV	KAV (1.0)
KAE	VAE (1.0)
VAE	KAA (1.0)
KAV	VAV (0.5), KBBd (0.5)

3.3 Graphical Representation

Figure 1 visualizes the grammar as a directed graph with probabilities.

```
               KBG
              /   \
           0.667   0.333
            /       \
          VBG       VBBd
           |          |
          1.0         |
           |          |
          KBBd        |
          / | \        |
      0.667|  |0.167   |
        /  |  \        |
      VBBd |   VBA     |
       |   |    |      |
       |   |  0.5|     |
       |   |    |      |
       |   |   KBBd    |
       |   |   / \     |
       |   |  ... ...  |
       |   |           |
       \---/-----------/
          |
         VAA
        /   \
     0.857   0.143
      /       \
    KAA       KAV
    / \        |
 0.75 0.25     |
  /    \       |
VAV    VBG     |
 |             |
KAV           VAV
 |             |
 \_____________/
       |
      END
```

Figure 1: Graphical representation of the optimized grammar. Edge weights represent transition probabilities.

4 Interpretive Analysis of Probabilities

4.1 Constitutive Rules (Structural Necessities)

Some transitions have probability 1.0, indicating that they are not merely statistical regularities but *constitutive rules* of the interaction format:

- **VBG → KBBd (1.0)**: A seller greeting must be followed by a customer need articulation. This is a structural property of sales conversations. Without this transition, the conversation would lack its transactional purpose.
- **KAE → VAE (1.0)**: A customer inquiry must be answered by seller information. This reflects the normative expectation of reciprocity and cooperation.
- **VAE → KAA (1.0)**: Seller information is always followed by customer completion. This suggests that information exchange in market conversations is tightly coupled to transaction closure.
- **VAV → KAV (1.0)**: Seller farewells are always reciprocated by customer farewells. This is a ritualized closing sequence that marks the end of the interaction.

4.2 Statistical Regularities (Empirical Contingencies)

Other transitions have probabilities between 0 and 1, reflecting empirical frequencies that could vary across contexts:

- **KBG → VBG (0.667)** vs. **KBG → VBBd (0.333)**: In two-thirds of cases, a customer greeting is reciprocated; in one-third, the seller responds directly with an inquiry (skipping the reciprocal greeting). This shows that the reciprocal greeting is the norm but not obligatory.
- **KBBd → VBBd (0.667)** vs. **KBBd → VAA (0.167)** vs. **KBBd → VBA (0.167)**: Most customer needs are followed by seller inquiries, but sometimes directly by completion (immediate purchase) or reaction (consultative response).
- **VBBd → KBA (0.444)** vs. **VBBd → VAA (0.222)** vs. **VBBd → KBBd (0.222)** vs. **VBBd → KAA (0.111)**: Seller inquiries have the most varied outcomes—responses, completions, need loops (upselling), or customer completions (early exit).
- **KBA → VBA (0.5)** vs. **KBA → VAA (0.5)**: Customer responses are

equally likely to lead to seller reactions or direct completions. This suggests a strategic choice point.

- **VAA → KAA (0.857)** vs. **VAA → KAV (0.143)**: Most seller completions are followed by customer completions; rarely, directly by customer farewell (abbreviated closing).
- **KAA → VAV (0.75)** vs. **KAA → VBG (0.25)**: Three quarters of customer completions lead to seller farewells; one quarter lead to a restart of the conversation (new greeting). The restart option occurs when a new customer arrives or when the same customer makes an additional purchase.

4.3 The Upselling Loop

The transition **VBA → KBBd (0.5)** deserves special attention. This loop—from seller reaction back to customer need—is the grammatical representation of *upselling*. When a seller says "Anything else?" (VBA), the customer often responds with an additional need (KBBd).

Crucially, this loop occurs in only half of the cases (0.5). The other half lead to completion (VAA, 0.25) or inquiry (KAE, 0.25). This suggests that upselling is neither mandatory nor rare—it is a strategic option that sellers can deploy, and customers can accept or deflect.

4.4 The Restart Option

The transition **KAA → VBG (0.25)** is particularly interesting. A customer completion (KAA) is normally followed by farewell (VAV, 0.75), but in a quarter of cases, it is followed by a seller greeting (VBG). This indicates that conversations can restart after completion—for example, when a new customer arrives (as in T2 and T5) or when the same customer makes an additional purchase.

4.5 The New Customer Arrival

The transition **KAV → KBBd (0.5)** shows that a customer farewell can be "cancelled" when a new customer arrives. In T5, a farewell (KAV) is immediately followed by a new customer need (KBBd). This is the only transition that breaks the otherwise strict sequencing of phases.

5 Constitutive Rules vs. Statistical Regularities

5.1 The Distinction

The ARS framework maintains a strict separation between two levels of description:

1. **Constitutive rules**: Structural constraints that define what counts as a well-formed sequence. These are binary (a sequence either conforms or it does not). In the grammar, transitions with probability 1.0 in the corpus are candidates for constitutive rules.

2. **Statistical regularities**: Empirical frequencies of transitions. These are probabilistic and can be updated as new data arrives. They reflect what *happens* in the data, not what *must* happen.

Table 4: Constitutive rules identified from the corpus

Rule	Probability	Interpretation
VBG → KBBd	1.0	Seller greeting must be followed by customer need
KAE → VAE	1.0	Customer inquiry must be answered
VAE → KAA	1.0	Information must lead to completion
VAV → KAV	1.0	Farewells are reciprocal

5.2 From Statistics to Constitutivity

A statistical regularity can become a constitutive rule through methodological decision. For example, the transition `VBG → KBBd (1.0)` never varied in the corpus. We could treat it as a constitutive rule: "In sales conversations, a seller greeting must be followed by a customer need articulation." This turns an empirical observation into a normative constraint.

However, this decision is not automatic. It requires methodological reflection: Is this truly a necessary feature of the interaction format, or could it vary in other contexts? The ARS approach is to treat transitions as statistical until counterexamples force a revision of the structural grammar.

5.3 The Methodological Significance

This flexibility is crucial for XAI and neuro-symbolic integration. It allows the analyst to:

1. Start with purely statistical learning (no prior constraints).
2. Identify transitions that never vary in the data.
3. Elevate them to constitutive rules after methodological reflection.
4. Use the resulting hybrid system for prediction and explanation.

The system thus moves from purely empirical pattern recognition (System 1) to rule-based explanation (System 2)—a shift that mirrors Kahneman's distinction between fast and slow thinking (Kahneman, 2011).

6 Toward Neuro-Symbolic Implementation

6.1 The Dual-Dynamics Architecture

The ARS grammar naturally suggests a dual-dynamics architecture for neuro-symbolic implementation:

1. **Symbolic component** (System 2): The grammar rules, including both constitutive rules and statistical probabilities. This component is inspectable, falsifiable, and explainable.
2. **Neural component** (System 1): A neural network that learns transition probabilities from data and predicts next symbols. This component is fast, pattern-based, and scalable.
3. **Hybrid integration**: The neural component provides fast predictions; the symbolic component validates them against constitutive rules and provides explanations.

[caption=Pseudocode for neuro-symbolic ARS] Pseudocode: ARS 5.0 Dual-Dynamics Architecture

class ARSNeuroSymbolicSystem:

def $_{init_{(self):Symboliccomponentself.grammar=ARSGrammar()PCFGwithprobabilitiesself.counts=zero_matrix(12,12)Transitioncounts}}$

Neural component self.neural$_n etwork = NeuralNetwork(input_dim = 12, hidden = 64, output_dim = 12)$

Constitutive rules (hard constraints) self.constitutive$_r ules = (KBG, VBG) : True, (VBG, KBBd) : T$

def update$_{s}ymbolic(self, from_{s}ym, to_{s}ym)$: """$Fast, counting-basedupdate(System2)$"""$self.coun$
$1row_{s}um = sum(self.counts[from_{s}ym])self.grammar.probs[from_{s}ym] = self.counts[from_{s}ym]/r$

def update$_{n}eural(self, from_{s}ym, to_{s}ym)$: """$Slow, gradient-basedupdate(System1)$"""$loss =$
$cross_{e}ntropy(self.neural_{n}etwork(from_{s}ym), to_{s}ym)loss.backward()optimizer.step()$

def predict$_{n}ext(self, from_{s}ym)$: """$Hybridprediction$"""$neural_{p}robs = self.neural_{n}etwork(from_{s}$
$self.grammar.probs[from_{s}ym]return0.5 * neural_{p}robs + 0.5 * symbolic_{p}robs$

def is$_{v}alid(self, from_{s}ym, to_{s}ym)$: """$Checkconstitutiverules$"""$returnself.constitutive_{r}ules.get(($

6.2 Explainability Through Proof Trees

The symbolic component enables explainability through proof trees. For any well-formed sequence, the grammar can produce a derivation:

[caption=Proof tree for a well-formed sequence] well$_{f}ormed([KBG, VBG, KBBd])$¸$start(KBG)[1.0]$¸
0.667Ö1.0Ö0.0 = 0.0

This proof tree makes the reasoning transparent. Each step is justified by a rule, and each rule has an explicit probability.

6.3 Implementation Options

The ARS grammar can be implemented in various neuro-symbolic frameworks:

Table 5: Implementation options for ARS 5.0

Framework	**Language**	**Best for**
DeepProbLog	Prolog/Python	Research, explainability
PyTorch	Python	Flexibility, prototyping
Flux.jl	Julia	Scientific computing, performance
Candle	Rust	Production, edge computing

Each framework has its own strengths, but all can implement the same dual-dynamics architecture.

7 Conclusion

This paper has reconstructed the empirical grammar of eight market conversations, from raw transcripts to an empirically optimized probabilistic grammar. The analysis yielded three main insights:

1. **Empirical**: The optimized grammar achieves high correlation with the data ($r = 0.925$) and reveals both constitutive rules (transitions with probability 1.0) and statistical regularities.

2. **Interpretive**: The upselling loop (VBA $\rightarrow$ KBBd, 0.5) and restart option (KAA $\rightarrow$ VBG, 0.25) highlight strategic choices in sales interactions. The farewell-cancellation (KAV $\rightarrow$ KBBd, 0.5) shows how new customers can enter ongoing interactions.

3. **Methodological**: The distinction between constitutive rules and statistical regularities provides a foundation for explainable, falsifiable, and adaptable neuro-symbolic systems.

The grammar presented here is not a static artifact. It can be updated as new data arrives (statistical plasticity) and revised when structural anomalies are detected (structural stability). In this sense, it is a living neuro-symbolic model—an empirical grammar that learns while remaining explainable.

References

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Koop, P. (1994). *Grammatikinduktion empirisch gesicherter Verkaufsgespräche*. Scheme source code.

Koop, P. (2026). *From Scheme to DeepProbLog: ARS as a Methodological Blueprint for Modern Neuro-Symbolic Programming*. the-last-freedom.org.

Manhaeve, R., Dumancic, S., Kimmig, A., Demeester, T., & De Raedt, L. (2018). DeepProbLog: Neural probabilistic logic programming. *Advances in Neural Information Processing Systems*, 31.

Oevermann, U., Allert, T., Konau, E., & Krambeck, J. (1979). The methodology of objective hermeneutics. In H.-G. Soeffner (Ed.), *Interpretative Procedures in the Social and Text Sciences* (pp. 352-434). Metzler.