

Algorithmic Recursive Sequence Analysis 2.0 – Example with Critical Reliability Assessment

Paul Koop

November 2024

post@paul-koop.org

Abstract:

The method of Algorithmic Recursive Sequence Analysis is applied to the transcribed protocol of a sales conversation. It is assumed that in order to reconstruct the speakers' intentions, one must first clarify what the interactions could mean in terms of rules of a probabilistic context-free grammar. After a preliminary investigation to form hypotheses about sales conversations, the interactions are examined — oriented on but independent of the hypothetically formed grammar — such that a terminal string is derived from the interactions, assigning a terminal symbol to each interact. Subsequently, a grammar is induced from the terminal string. This empirical terminal string is parsed for well-formedness with respect to the induced grammar. Artificial terminal strings are then generated from this grammar and tested for significant agreement with the empirically obtained grammar. In a further step, the grammar can be handed over to a (hybrid) multi-agent system that orients its simulation of sales conversations to the rules and protocol of the grammar.

Reliability of Coding: Critical Examination and Adjustment

The basis of any further analysis is the assignment of 59 inter-acts to 12 categories by two independent coders (Coder 1 and Coder 2). The originally used formula

$$R_{\text{ars}} = \frac{N \cdot Z}{I} = \frac{2 \cdot 35}{118} = 0.59$$

measures the observed agreement ($p_o = 0.59$) but does **not** correct for the proportion of agreement expected by chance. With 12 equally likely categories, the chance expectation is

$$p_e = \frac{1}{12} \approx 0.0833.$$

The corrected coefficient (Cohen's Kappa) is therefore

$$\kappa = \frac{p_o - p_e}{1 - p_e} = \frac{0.59 - 0.0833}{1 - 0.0833} \approx 0.55.$$

With $\kappa \approx 0.55$, the adjusted reliability lies only in the lower range of "moderate" (according to Landis & Koch) and **falls short** of the usual acceptance thresholds

of $\kappa \geq 0.70$ (exploratory) or $\kappa \geq 0.80$ (confirmatory). The original formula systematically overestimates the true agreement. For robust content analysis, a reliability of 0.55 is insufficient.

Cross-Tabulations and Additional Chi-Square Tests

The Chi-square values refer to **goodness-of-fit tests for each coder individually**, not to the agreement between coders. They test whether the observed distribution of categories deviates from an expected distribution.

Table 1: Chi-square tests per coder

Coder	χ^2	df	Asymp. Sig.	pre-set α
1	2.6	6	0.86	0.9
2	2.0	5	0.85	0.9

The high asymptotic significances ($p > 0.05$) indicate that for neither coder is there a significant deviation from the assumed base distribution. However, these tests have **no bearing on the inter-coder reliability**. They are at best relevant for the validity of each individual coder with respect to an external reference distribution.

Conclusion on Statistics

- The original reliability value ($R_{ars} = 0.59$) is **misleading** because it lacks a chance correction.
- The adjusted Kappa value of **0.55** is **insufficient** for robust content analysis.
- The Chi-square tests do not compensate for the lack of agreement between coders.
- For a methodologically sound study, an **improvement of the coding scheme** or **re-training of the coders** would be necessary before starting grammar induction.

Processing: Hypothesis Formation and Preliminary Investigation

The procedure described here corresponds to a methodological approach that combines qualitative social research with formal grammar theory to investigate the structure and interaction dynamics of sales conversations. The analysis is divided into several steps that allow a systematic examination of the conversational structure and can ultimately lead to a simulation by a multi-agent system. Here is an overview and detailed consideration of the individual steps:

1. **Hypothesis formation and preliminary investigation:**

- First, a hypothesis about the structure of sales conversations is formed. The preliminary assumption (grammar) states that sales conversations are divided into greeting, sales section, and farewell.
- This assumption is used for orientation, but the interactions are to be analyzed independently in order to draw inductive conclusions.

2. Analysis of interactions and assignment of terminal symbols:

- The individual interaction components of the conversation are translated into a "terminal string" by assigning a symbol to each conversational act (e.g., greeting, needs clarification).
- By assigning terminal symbols to each interact (e.g., KBG for the customer's greeting), a sequential sequence of symbols is created that maps the conversation in a symbolic form.

3. Induction of a grammar:

- A grammar is induced from the resulting string, describing the structure of the interactions. This grammar attempts to capture the recurrent patterns and transitions in the conversation.
- The well-formedness of the empirically observed terminal string is checked by parsing it against the induced grammar. This determines whether the interaction sequences match the generated structure.

4. Generation and comparison of artificial strings:

- Using the induced grammar, new artificial terminal strings are generated that simulate the structure of the sales conversation.
- These artificial conversations are examined for significant agreement with the original, empirically obtained grammar to check their consistency and representative accuracy.

5. Simulation by a multi-agent system:

- In the final step, the induced grammar is implemented in a multi-agent system that can simulate sales conversations. The agents follow the rules defined by the grammar and interact according to the conversation protocol.
- This system can be used for hypothesis testing or for analyzing possible variations in sales conversations.

Summary of the Preliminary Grammar

The sales conversation (VKG) is defined as a structure consisting of the following elements:

- **Greeting (BG):** Greeting by customer (KBG) and seller (VBG).

- **Sales section (VT):** Needs section (B) and closing section (A).
 - Needs section includes needs clarification (BBd) with customer statements (KBBd) and seller responses (VBBd) as well as needs argumentation (BA) with arguments from both sides (KBA, VBA).
 - Closing section contains objections (AE) and sales closing (AA) with objections and closing arguments from both sides (KAE, VAE, KAA, VAA).
- **Farewell (AV):** Farewell by customer (KAV) and seller (VAV).

The terminal symbols used as "end elements" of the grammar are:

- **Greeting and farewell:** KBG, VBG, KAV, VAV.
- **Needs section and arguments:** KBBd, VBBd, KBA, VBA.
- **Closing section and objections:** KAE, VAE, KAA, VAA.

This methodological approach enables a precise analysis and modeling of the structures and dynamics that occur in sales conversations. It provides insights into the social and linguistic rules that determine such interactions and allows the simulation of the findings for further investigation.

Task: Analysis of Interactions

Analyze the interactions of the following transcript of a recorded sales conversation for their possible meanings. Exclude inappropriate meanings through the readings of the preceding interact and assign a category and a terminal symbol to each interact:

— Beginning of Text4 Market, 11:00 a.m.

(Aachen, 06/28/94, vegetable stall)

(unintelligible)

4. C1.1: Listen, I'll take some mushrooms.
5. S1.1: Eh, brown ones?
6. C1.2: No, light ones.
7. S1.2: Light ones.
8. C1.3: Mhmm.
9. (unintelligible)
10. C1.4: You don't see them.
11. S1.3: Yes, it doesn't matter, they're both fresh.
12. C1.5: Or, what about, about, eh...

13. S1.4: You can let them sit longer.
14. C1.6: No, but chanterelles.
15. S1.5: Ah, they're great.
16. (unintelligible)
17. C1.7: Can I put them in rice salad?
18. (unintelligible)
19. C1.8: I don't need to boil them or something?
20. S1.6: Ehh, raw, but you have to put them in the pan a bit.
21. C1.9: I will.
22. S1.7: A little bit.
23. (unintelligible)
24. C1.10: But I can put them in rice salad.
25. S1.8: Yes, that's no problem, you just have to...
26. C1.11: A little, yes.
27. S1.9: A little in eh, not like the mushrooms, I put them in raw too.
28. C1.11: Eh yes.
29. S1.10: A hundred, right?
30. C1.12: Yes please. What else can I get here?
31. S1.11: Wild berries? I have them too.
32. C1.13: (unintelligible) What about strawberries?
33. (unintelligible)
34. C1.14: What else do you have?
35. S1.12: Don't you feel like raspberries? Or currants, I have them too.
36. C1.15: Yes. (Pause) We'll take one of each.
37. S1.13: Currants are a pound, you can keep them longer too.
38. (Pause, unintelligible)
39. C1.16: Then I have, I think, everything for home.
40. S1.14: Yes, until the day after tomorrow, right.
41. C1.17: No. (Pause) Look, the young man has to provide for you.
42. S1.15: Yes, so that our, our, eh...
43. C1.18: Yes.

44. S1.16: (unintelligible) So that our brain doesn't start cooking.
45. C1.19: So.
46. S1.17: Sixty, four marks sixty, eight marks sixty, twelve marks fifty.
47. C1.20: You can't close the door.
48. S1.18: Twelve marks fifty. (Pause) I know, today I'm outrageous again...
49. C1.21: Yes.
50. S1.19: But, but, but, one mark (unintelligible) more.
51. C1.22: Listen.
52. S1.20: What do you need?
53. C1.23: To the eh Barmer, but I'll come, I'll come later, then I'll put it under.
54. S1.21: No, otherwise leave it here. Thirteen, fifteen, twenty marks.
55. C1.24: Thank you.
56. S1.22: Until the day after tomorrow.
57. C1.25: Thank you very much.
58. S1.23: Yes. End of Text4

Processing: Coding of Interactions

To analyze the interactions in this sales conversation and assign a terminal symbol to each interact, we consider how the contributions to the conversation contribute to the assumed categories (greeting, needs section, needs argumentation, sales closing, farewell). By sequencing the interactions and excluding inappropriate meanings, a clear structure emerges. Here is the analysis with assigned categories and terminal symbols:

Interactions 4–10:

- **4. C1.1: "Listen, I'll take some mushrooms."** – Customer's statement of need (needs clarification) → KBBd
- **5. S1.1: "Eh, brown ones?"** – Seller clarifies need (needs clarification) → VBBd
- **6. C1.2: "No, light ones."** – Customer specifies need (needs clarification) → KBBd
- **7. S1.2: "Light ones."** – Seller confirms need (needs clarification) → VBBd
- **8. C1.3: "Mhmh."** – Customer agreement, no new need → KBBd

- **10. C1.4: "You don't see them."** – Customer clarifies need, possible uncertainty → KBBd

Interactions 11–15:

- **11. S1.3: "Yes, it doesn't matter, they're both fresh."** – Seller argumentation, excludes alternatives (needs argumentation) → VBA
- **12. C1.5: "Or, what about, about, eh..."** – Customer checks alternatives (needs clarification) → KBBd
- **13. S1.4: "You can let them sit longer."** – Seller argues for product (needs argumentation) → VBA
- **14. C1.6: "No, but chanterelles."** – Customer expresses interest in alternative product (needs clarification) → KBBd
- **15. S1.5: "Ah, they're great."** – Seller praises alternative product (needs argumentation) → VBA

Interactions 17–24:

- **17. C1.7: "Can I put them in rice salad?"** – Customer asks about use (needs argumentation) → KBA
- **19. C1.8: "I don't need to boil them or something?"** – Customer asks about preparation (needs argumentation) → KBA
- **20. S1.6: "Ehh, raw, but you have to put them in the pan a bit."** – Seller gives preparation advice (needs argumentation) → VBA
- **21. C1.9: "I will."** – Customer agrees (needs argumentation) → KBA
- **24. C1.10: "But I can put them in rice salad."** – Customer confirms use option (needs argumentation) → KBA
- **25. S1.8: "Yes, that's no problem, you just have to..."** – Seller confirms (needs argumentation) → VBA

Interactions 29–32:

- **29. S1.10: "A hundred, right?"** – Seller's price offer (sales closing) → VAA
- **30. C1.12: "Yes please. What else can I get here?"** – Customer accepts and asks for more products (sales closing) → KAA
- **31. S1.11: "Wild berries? I have them too."** – Seller offers additional product (needs argumentation) → VBA
- **32. C1.13: "(unintelligible) What about strawberries?"** – Customer interested in alternatives (needs clarification) → KBBd

Interactions 34–39:

- **34. C1.14: "What else do you have?"** – Inquiry about further options (needs clarification) → KBBd
- **35. S1.12: "Don't you feel like raspberries? Or currants, I have them too."** – Seller offers alternatives (needs argumentation) → VBA
- **36. C1.15: "Yes. (Pause) We'll take one of each."** – Customer agrees, purchase decision (sales closing) → KAA
- **37. S1.13: "Currants are a pound, you can keep them longer too."** – Seller adds information (needs argumentation) → VBA
- **39. C1.16: "Then I have, I think, everything for home."** – Customer signals completion of purchase (sales closing) → KAA

Interactions 40–58:

- **40. S1.14: "Yes, until the day after tomorrow, right."** – Seller says goodbye (farewell) → VAV
- **41. C1.17: "No. (Pause) Look, the young man has to provide for you."** – Casual conversation, no new meaning
- **46. S1.17: "Sixty, four marks sixty, eight marks sixty, twelve marks fifty."** – Seller calculates total (sales closing) → VAA
- **49. C1.21: "Yes."** – Customer accepts price (sales closing) → KAA
- **54. C1.24: "Thank you."** – Customer says goodbye (farewell) → KAV
- **56. S1.22: "Until the day after tomorrow."** – Seller says goodbye again (farewell) → VAV
- **57. C1.25: "Thank you very much."** – Customer repeats farewell (farewell) → KAV

In summary, the interactions can be assigned to the categories greeting (none explicit in the transcript), needs clarification, needs argumentation, sales closing, and farewell, with each category assigned the corresponding terminal symbols.

Task: Adding the Greeting

Assume the missing greeting by the customer and the seller and provide the complete terminal string for which a grammar is to be induced.

Processing: Complete Terminal String

To construct a complete terminal string that also accounts for the missing greeting, we can add a greeting by the customer (KBG) and the seller (VBG) at the beginning of the conversation. This provides a more logical and complete sequence that can serve as a basis for grammar induction.

The complete terminal string is as follows:

**KBG VBG KBBd VBBd KBBd VBBd KBBd VBA KBBd VBA
KBBd VBA KBA KBA VBA KBA KBA VBA VAA KAA VBA
KBBd KBBd VBA KAA VBA KAA VAA KAA VAV KAV VAA
KAA KAV VAV KAV**

Here is the legend for the categories:

- **KBG**: Greeting by customer
- **VBG**: Greeting by seller
- **KBBd**: Needs clarification by customer
- **VBBd**: Needs clarification by seller
- **VBA**: Needs argumentation by seller
- **KBA**: Needs argumentation by customer
- **VAA**: Sales closing by seller
- **KAA**: Sales closing by customer
- **VAV**: Farewell by seller
- **KAV**: Farewell by customer

This sequence provides a clear structure of the interactions that can be used for grammar induction.

Task: Inducing a Probabilistic Grammar

Generate a probabilistic grammar that describes the following terminal string:
KBG VBG KBBd VBBd KBBd VBBd KBBd VBA KBBd VBA KBBd VBA
KBA KBA VBA KBA KBA VBA VAA KAA VBA KBBd KBBd VBA KAA
VBA KAA VAA KAA VAV KAV VAA KAA KAV VAV KAV

1. **Structural analysis**: Analyze the string for recurring patterns or phrase structures and identify potential rules and hierarchies that could describe the string.
2. **Probability distribution**: For each rule, create probabilities based on the frequency of individual elements and transitions in the string.
3. **Combine rules and probabilities**: Formulate the rules of the grammar and assign probabilities to them so that the grammar describes the given string as precisely as possible.
4. **Output format**: Present the induced probabilistic grammar in BNF format, including the associated probabilities for each rule.

Processing: Induced Probabilistic Grammar

Several steps are required to convert the given string into a probabilistic grammar. The goal is to identify recurring patterns and define rules with probabilities that reflect the observed transitions. The following describes the steps for creating a probabilistic context-free grammar (PCFG) for the given terminal string.

1. Structural Analysis

The string can be divided into logical segments that often contain recurring elements. A typical structure of a sales conversation with greeting, needs clarification, needs argumentation, sales closing, and farewell can be identified as follows:

- **Greeting:** KBG VBG
- **Needs clarification and argumentation:** Repeated interplay between KBBd, VBBd, KBA, VBA
- **Sales closing:** KAA, VAA
- **Farewell:** KAV, VAV

2. Identification of Rules and Probability Distribution

Based on the frequency of certain combinations in the sequence, we can derive transition probabilities. We create a rule for each observation and calculate probabilities based on the frequency of transitions.

3. Formulation of Rules and Probabilities

We define the probabilistic grammar in a Backus-Naur Form (BNF) style with probabilities for the rules. The probabilities are based on the frequency of the respective rule in the given string.

4. Probabilistic Grammar in BNF Format

The resulting PCFG that describes the structure and probabilities of the transitions is as follows:

Start Symbol $\langle \text{Start} \rangle ::= \langle \text{Greeting} \rangle \langle \text{Needs} \rangle \langle \text{Closing} \rangle \langle \text{Farewell} \rangle$

Greeting $\langle \text{Greeting} \rangle ::= \text{KBG VBG} [1.0]$

Needs The needs section consists of several segments with repeating patterns:

$\langle \text{Needs} \rangle ::= \langle \text{NeedsSegment} \rangle \langle \text{Needs} \rangle [0.8] \mid \langle \text{NeedsSegment} \rangle [0.2]$

$\langle \text{NeedsSegment} \rangle ::= \text{KBBd VBBd} [0.4] \mid \text{KBBd VBA} [0.3] \mid \text{KBA VBA} [0.3]$

Closing <Closing> ::= KAA VAA [0.6] | VAA KAA [0.4]

Farewell <Farewell> ::= KAV VAV [0.7] | VAV KAV [0.3]

Consolidated Grammar

<Start> ::= <Greeting> <Needs> <Closing> <Farewell>

<Greeting> ::= KBG VBG [1.0]

<Needs> ::= <NeedsSegment> <Needs> [0.8] | <NeedsSegment> [0.2]

<NeedsSegment> ::= KBBd VBBd [0.4] | KBBd VBA [0.3] | KBA VBA [0.3]

<Closing> ::= KAA VAA [0.6] | VAA KAA [0.4]

<Farewell> ::= KAV VAV [0.7] | VAV KAV [0.3]

Explanation of Steps and Probabilities

- **Greeting:** The greeting occurs once at the beginning of the conversation and is unambiguous, so it has a probability of 1.0.
- **Needs:** The needs section consists of several sequences (segments) that can follow each other repeatedly and variably, hence a recursion step with probability 0.8 for repetition and 0.2 for ending the needs section.
- **Closing:** The closing has two common patterns (KAA VAA and VAA KAA) with different probabilities, based on their frequency in the sequence.
- **Farewell:** The farewell shows preferred sequences, with KAV VAV being more frequent.

Conclusion

This probabilistic grammar describes the sales conversation in such a way that the probabilities capture different variations that still correspond to the observed sequence structure. The generated PCFG also allows the generation of new, probable conversation sequences based on the original structure. However, the caveat of insufficient inter-coder reliability ($\kappa \approx 0.55$) remains. The grammar is therefore to be understood as a preliminary explication, not as an empirically validated reconstruction.