NLP Strategies Evaluation

Summary

Extraction Strategies	3
Intent Strategies	1
Test Sentences	10000
Timestamp	2026-02-12 16:46:26

Station Extraction Evaluation

RULE_BASED Extractor

Classification Metrics

Overall Accuracy	70.09% (7009/10000)
Precision (CORRECT)	72.02%
Recall (CORRECT)	81.35%
F1 Score (CORRECT)	76.40%

Per-Category Accuracy

Category	Accuracy
CORRECT	81.35%
NOT_FRENCH	19.85%
NOT_TRIP	93.92%

Extraction Breakdown

Complete Extractions	67.23% (6723)
Partial Extractions	0.00% (0)

Errors	32.77% (3277)

Timing

Mean	16.97 ms
Median	16.62 ms
P95	18.50 ms
Min	15.31 ms
Max	42.95 ms
Std Dev	1.64 ms

HF_NER Extractor

Classification Metrics

Overall Accuracy	52.35% (5235/10000)
Precision (CORRECT)	72.14%
Recall (CORRECT)	32.49%
F1 Score (CORRECT)	44.80%

Per-Category Accuracy

Category	Accuracy
CORRECT	32.49%
NOT_FRENCH	66.15%
NOT_TRIP	100.00%

Extraction Breakdown

Complete Extractions	26.81% (2681)
Partial Extractions	0.00% (0)
Errors	73.19% (7319)

Mean	34.04 ms
Median	34.49 ms
P95	41.51 ms
Min	16.36 ms
Max	1973.50 ms
Std Dev	20.34 ms

SPACY Extractor

Classification Metrics

Overall Accuracy	54.93% (5493/10000)
Precision (CORRECT)	93.08%
Recall (CORRECT)	26.23%
F1 Score (CORRECT)	40.92%

Per-Category Accuracy

Category	Accuracy
CORRECT	26.23%
NOT_FRENCH	94.74%
NOT_TRIP	100.00%

Extraction Breakdown

Complete Extractions	16.77% (1677)
Partial Extractions	0.00% (0)
Errors	83.23% (8323)

Timing

Mean	6.62 ms
Median	6.29 ms
P95	9.37 ms
Min	3.52 ms
Max	32.54 ms
Std Dev	1.47 ms

Extraction Strategy Comparison

Metric	RULE_BASED	HF_NER	SPACY
Accuracy	70.1%	52.3%	54.9%
Precision	72.0%	72.1%	93.1%
Recall	81.4%	32.5%	26.2%
F1	76.4%	44.8%	40.9%
Mean Time	17.0 ms	34.0 ms	6.6 ms
Median Time	16.6 ms	34.5 ms	6.3 ms
P95 Time	18.5 ms	41.5 ms	9.4 ms

Intent Classification Evaluation

RULE_BASED Classifier

Overall Metrics

Accuracy	66.13% (6613/10000)
Macro F1	63.56%
Weighted F1	69.28%

Per-Class Metrics

Class	Precision	Recall	F1	Support
NOT_FRENCH	84.9%	50.1%	63.0%	2207
NOT_TRIP	35.1%	88.0%	50.2%	1841
TRIP	95.4%	65.3%	77.5%	5952

Confusion Matrix

Expected \ Predicted	NOT_FRENCH	NOT_TRIP	TRIP
NOT_FRENCH	1105	1102	0

NOT_TRIP	32	1621	188
TRIP	165	1900	3887

Timing

Mean	3.94 ms
Median	2.57 ms
P95	10.83 ms
Min	0.00 ms
Max	1165.68 ms
Std Dev	12.17 ms