

# NLP Strategies Evaluation

## Summary

Extraction Strategies	3
Intent Strategies	1
Test Sentences	10000
Timestamp	2026-02-13 11:03:42

## Station Extraction Evaluation

### *RULE\_BASED Extractor*

#### *Classification Metrics*

Overall Accuracy	69.93% (6993/10000)
Precision (CORRECT)	71.93%
Recall (CORRECT)	80.80%
F1 Score (CORRECT)	76.11%

#### *Per-Category Accuracy*

Category	Accuracy
CORRECT	80.80%
NOT_FRENCH	20.42%
NOT_TRIP	94.24%

#### *Extraction Breakdown*

Complete Extractions	66.58% (6658)
Partial Extractions	0.00% (0)

Errors	33.42% (3342)
--------	---------------

*Timing*

Mean	12.30 ms
Median	10.36 ms
P95	34.16 ms
Min	9.08 ms
Max	78.66 ms
Std Dev	7.61 ms

*HF\_NER Extractor*

*Classification Metrics*

Overall Accuracy	52.69% (5269/10000)
Precision (CORRECT)	72.30%
Recall (CORRECT)	32.71%
F1 Score (CORRECT)	45.05%

*Per-Category Accuracy*

Category	Accuracy
CORRECT	32.71%
NOT_FRENCH	66.44%
NOT_TRIP	100.00%

*Extraction Breakdown*

Complete Extractions	26.82% (2682)
Partial Extractions	0.00% (0)
Errors	73.18% (7318)

### *Timing*

<b>Mean</b>	28.95 ms
<b>Median</b>	23.29 ms
<b>P95</b>	52.41 ms
<b>Min</b>	15.12 ms
<b>Max</b>	3828.66 ms
<b>Std Dev</b>	39.75 ms

### *SPACY Extractor*

#### *Classification Metrics*

<b>Overall Accuracy</b>	54.16% (5416/10000)
<b>Precision (CORRECT)</b>	91.73%
<b>Recall (CORRECT)</b>	24.90%
<b>F1 Score (CORRECT)</b>	39.17%

#### *Per-Category Accuracy*

<b>Category</b>	<b>Accuracy</b>
CORRECT	24.90%
NOT_FRENCH	93.99%
NOT_TRIP	100.00%

#### *Extraction Breakdown*

<b>Complete Extractions</b>	16.09% (1609)
<b>Partial Extractions</b>	0.00% (0)
<b>Errors</b>	83.91% (8391)

### *Timing*

<b>Mean</b>	3.64 ms
<b>Median</b>	3.10 ms
<b>P95</b>	4.11 ms
<b>Min</b>	1.95 ms
<b>Max</b>	4471.13 ms
<b>Std Dev</b>	44.81 ms

## Extraction Strategy Comparison

Metric	RULE_BASED	HF_NER	SPACY
Accuracy	69.9%	52.7%	54.2%
Precision	71.9%	72.3%	91.7%
Recall	80.8%	32.7%	24.9%
F1	76.1%	45.0%	39.2%
Mean Time	12.3 ms	28.9 ms	3.6 ms
Median Time	10.4 ms	23.3 ms	3.1 ms
P95 Time	34.2 ms	52.4 ms	4.1 ms

## Intent Classification Evaluation

### ***RULE\_BASED Classifier***

#### *Overall Metrics*

Accuracy	66.04% (6604/10000)
Macro F1	63.25%
Weighted F1	69.02%

#### *Per-Class Metrics*

Class	Precision	Recall	F1	Support
NOT_FRENCH	81.9%	49.3%	61.6%	2214
NOT_TRIP	35.5%	87.8%	50.6%	1859
TRIP	95.2%	65.5%	77.6%	5927

#### *Confusion Matrix*

Expected \ Predicted	NOT_FRENCH	NOT_TRIP	TRIP
NOT_FRENCH	1092	1122	0

NOT_TRIP	31	1632	196
TRIP	210	1837	3880

*Timing*

Mean	2.58 ms
Median	1.74 ms
P95	7.21 ms
Min	0.00 ms
Max	171.98 ms
Std Dev	2.84 ms