

Useful visualization for evaluating systems which:

1. Should return only one interpretation
2. May decline to give an interpretation

Misunderstanding
wrong interpretation (error)

Trade-off

Non-understanding
failure to give an interpretation

How old are dinosaurs?



Misunderstanding leads to inappropriate response

We were first compiled just months ago

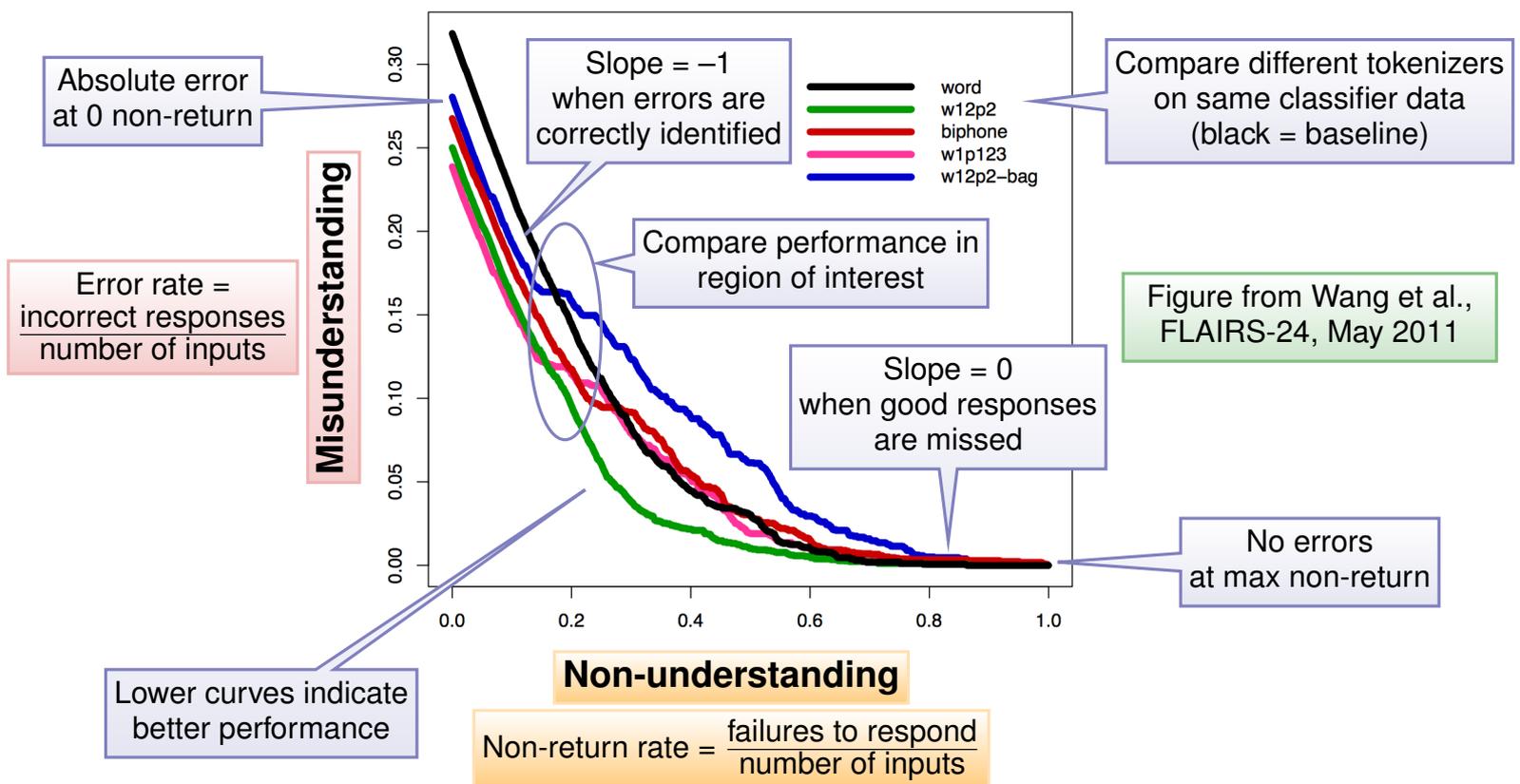
Could you repeat that?

Non-understanding leads to appropriate but uninformative response

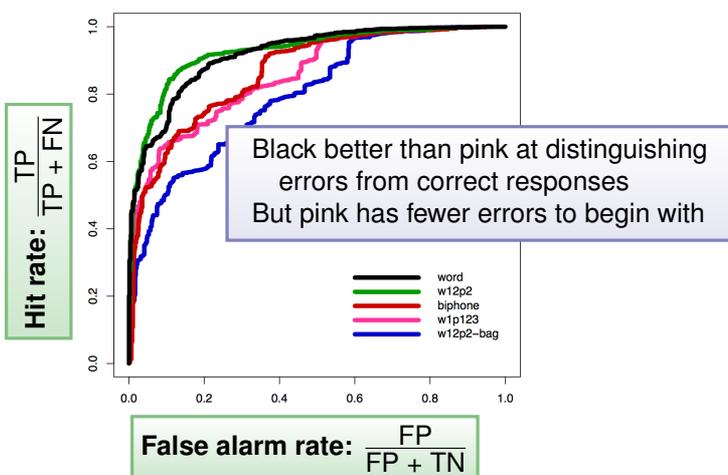
Non-understanding is the result of a 2-step process:

1. Determine the best interpretation (may turn out to be correct or incorrect)
2. Decide if it's likely to be correct (non-understanding when below a confidence threshold)

Visualization: Plot **misunderstanding** against **non-understanding** while varying the response threshold



ROC curves ignore actual error rate



TP = Correct response above threshold
 FP = Error response above threshold
 TN = Error response below threshold
 FN = Correct response below threshold

Precision/recall not very useful when each input receives at most one interpretation

Advantages of the error return plot

- Directly represents the tradeoff between misunderstanding and non-understanding.
- Visual indication if one system dominates another.
- Represents both the actual error rate, and the system's ability to identify errors.
- Slope of curve shows how much is gained by increasing non-understanding at each point.
- Shows performance at the appropriate level of non-understanding (varies by application).

The project or effort described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.