

How do we measure the difficulty of a dataset?

Premise	Animals are running
Hypothesis	Animals are outdoors
<hr/>	
Label	Entailment

→ **Hypo-only 72%**
Full-input 80% (+8%) → Dataset is “Easy”!

Failure of hypothesis-only model doesn’t mean dataset is difficult

Old Premise	Animals are running
New Premise	Entailment
Hypothesis	Animals are outdoors
<hr/>	
Label	Entailment

→ **Hypo-only 33%**
Perm-only 100%
Full-input 100% (+67%) → **Not Difficult!**

Label	Combinations				Premise	Ⓐ Animals are running
Entailment	Ⓐ+Ⓑ	Ⓒ+Ⓓ	Ⓔ+Ⓕ		Hypothesis	Ⓑ Animals are outdoors
Contradiction	Ⓐ+Ⓕ	Ⓒ+Ⓑ	Ⓔ+Ⓓ		<hr/>	
Neutral	Ⓐ+Ⓓ	Ⓒ+Ⓕ	Ⓔ+Ⓑ		Label	Entailment

Artifact exposes label by hiding
it in premise + hypothesis

→ **Hypo-only 33%**
Perm-only 33%
Full-input 100% (+67%) → **Not Difficult!**

In SNLI, “difficult” examples made easy with one extra word

Label	Premise	Hypothesis
Contradiction	A young boy hanging on a pole smiling at the <u>camera</u> .	The young boy is <u>crying</u> .
Contradiction	A boy smiles tentatively at the <u>camera</u> .	a boy is <u>crying</u> .
Contradiction	A happy child smiles at the <u>camera</u> .	The child is crying at the playground.
Contradiction	A girl shows a small child her <u>camera</u> .	A boy <u>crying</u> .
Entailment	A little boy with a baseball on his shirt is crying.	A boy is <u>crying</u> .
Entailment	Young boy crying in a stroller.	A boy is <u>crying</u> .
Entailment	A baby boy in overalls is crying.	A boy is <u>crying</u> .
Entailment	Little boy playing with his toy <u>train</u> .	A boy is playing with <u>toys</u> .
Entailment	A little boy is looking at a toy <u>train</u> .	A boy is looking at a <u>toy</u> .
Entailment	Little redheaded boy looking at a toy <u>train</u> .	A little boy is watching a toy <u>train</u> .
Entailment	A young girl in goggles riding on a toy <u>train</u> .	A girl rides a <u>toy train</u> .
Contradiction	A little girl is playing with tinker toys.	A little boy is playing with <u>toys</u> .
Contradiction	A toddler shovels a snowy driveway with a shovel.	A young child is playing with <u>toys</u> .
Contradiction	A boy playing with toys in a bedroom.	A boy is playing with toys at the <u>park</u> .

25% of all “difficult” examples are solvable with one premise word