Detecting Asymmetric Semantic Relations in Context

A Case Study on Hypernymy Detection

Yogarshi Vyas and Marine Carpuat

*SEM 2017 - 08/03/2017

Chess $\xrightarrow{\textbf{Hypernym?}}$ Game

Chess → Game ✓

Magnus Carlsen is the world **chess** champion

Hypernym?

The championship **game** was played in NYC.

Magnus Carlsen is the world **chess** champion

The championship **game** was played in NYC.

Magnus Carlsen is the world **chess** champion

Hypernym?

The poachers hunted the big **game**

Magnus Carlsen is the world **chess** champion

The poachers hunted the big **game**

Magnus Carlsen is the world **chess** champion

The championship **game** was played in NYC.

The poachers hunted the big **game**

Hypernymy Detection in Context

Input

Output

Hypernymy Detection in Context

Input

chess **game**

Output

Hypernymy Detection in Context

Input

Magnus Carlsen is the world **chess** champion

The championship **game** was played in NYC.

Output

Hypernymy Detection in Context

Input

Magnus Carlsen is the world **chess** champion

The championship **game** was played in NYC.

Output

Hypernymy Detection in Context

Input

Magnus Carlsen is the world **chess** champion

The championship **game** was played in NYC.

Output

✓ if game is a hypernym of chess in the given contexts

✗

Hypernymy Detection in Context

Input

Magnus Carlsen is the world **chess** champion

The championship **game** was played in NYC.

Output

✓ if game is a hypernym of chess in the given contexts

✗ otherwise

Motivation : Why Context?

- Benefit downstream tasks
  - Question Answering
  - Textual Inference

Motivation : Why Context?

- Benefit downstream tasks
  - Question Answering
  - Textual Inference
- Previous work relies on annotator to think about senses
  (Hearst, 1992; Zhitomirsky-Geffet and Dagan, 2009; Turney and Mohammad, 2013)

Previous Work

Previous Work : Context-PPDB

Shwartz and Dagan, *SEM 2016

Previous Work : Context-PPDB

- A dataset for **fine-grained lexical inference** in context
- 3750 examples, distributed over **6 semantic relation types**

Contributions

Contributions

- We construct **WHiC**, a new dataset of hypernym pairs in context
  - Automatically extracted from **WordNet**!

Contributions

- We construct **WHiC**, a new dataset of hypernym pairs in context
  - Automatically extracted from **WordNet**!
- WHiC enables
  - **Empirical comparisons** of context embeddings

Contributions

- We construct **WHiC**, a new dataset of hypernym pairs in context
  - Automatically extracted from **WordNet**!
- WHiC enables
  - **Empirical comparisons** of context embeddings
  - Analysis of **sensitivity to context,** and **directionality**

Contributions

- We construct **WHiC**, a new dataset of hypernym pairs in context
  - Automatically extracted from **WordNet**!
- WHiC enables
  - **Empirical comparisons** of context embeddings
  - Analysis of **sensitivity to context,** and **directionality**
- We provide experimental evidence of **challenging** nature of task

Outline

- WHiC : A Dataset for Hypernymy Detection in Context
  - Dataset Desiderata
  - Dataset Creation
- Two Models of Context Representation
- Experiments and Analysis

Outline

- WHiC : A Dataset for Hypernymy Detection in Context
  - Dataset Desiderata
  - Dataset Creation
- Two Models of Context Representation
- Experiments and Analysis

Dataset Desiderata

Dataset Desiderata

- Test sensitivity of models to changing contexts

Dataset Desiderata

- Test sensitivity of models to changing contexts
- Differentiate between **hypernymy** and **semantic similarity**
  - (Chess , Game) **v/s** (Chess , Checkers)

Dataset Desiderata

- Test sensitivity of models to changing contexts
- Differentiate between **hypernymy** and **semantic similarity**
  - (Chess , Game) **v/s** (Chess , Checkers)
- Large enough to train supervised models

WHiC : WordNet Hypernyms in Context

WHiC : WordNet Hypernyms in Context

...

Word : Representation
Example : the drawing was not a faithful representation

Word : Drawing, Art
Example : he did complicated pen-and-ink drawings

Word : Study, Sketch
Example : he made several studies before the final painting

WHiC : WordNet Hypernyms in Context

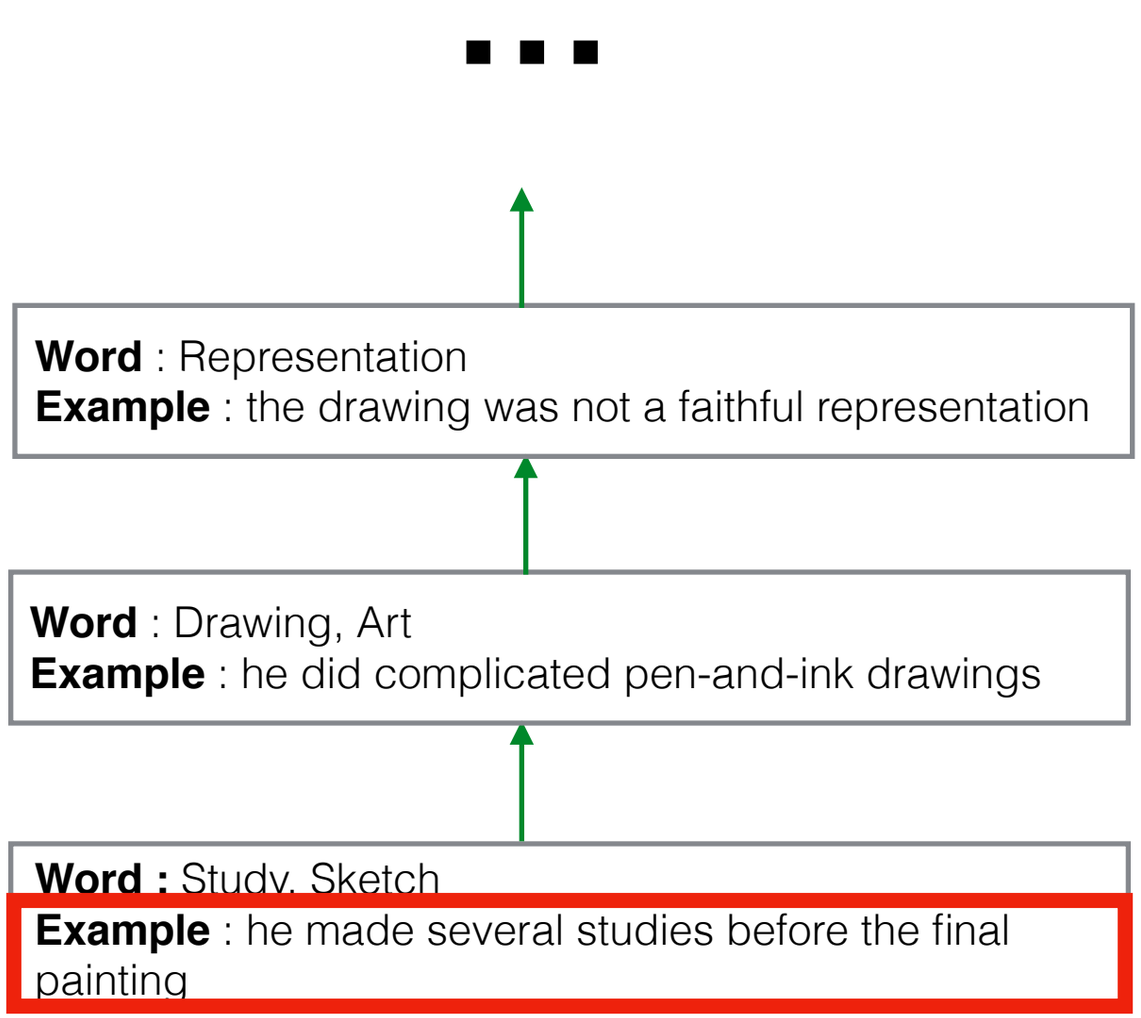
...

Word : Representation
Example : the drawing was not a faithful representation

Word : Drawing, Art
Example : he did complicated pen-and-ink drawings

Word : Study, Sketch
Example : he made several studies before the final painting

WHiC : WordNet Hypernyms in Context



WHiC : Positive Examples

Word : Study
Example : he knocked on the door of the study

Word : Study
Example : he made several studies before the final painting

w_1, w_2	c_1, c_2	IsHypernym(w_1, w_2) ?

WHiC : Positive Examples

Word : Room
Example : the rooms were small but comfortable

Word : Study
Example : he knocked on the door of the study

Word : Drawing
Example : he did complicated pen-and-ink drawings

Word : Study
Example : he made several studies before the final painting

w_1, w_2	c_1, c_2	IsHypernym(w_1, w_2) ?

WHiC : Positive Examples

Word : Room
Example : the rooms were small but comfortable

Word : Study
Example : he knocked on the door of the study

Word : Drawing
Example : he did complicated pen-and-ink drawings

Word : Study
Example : he made several studies before the final painting

w_1, w_2	c_1, c_2	IsHypernym(w_1, w_2) ?
Study, Room	He knocked on the door of the **study** The **rooms** were small but comfortable	✓
Study, Drawing	He made 2 **studies** before the painting He did pen-and-ink **drawings**	✓

WHiC : Negative Examples

Word : Room
Example : the rooms were small but comfortable

Word : Drawing
Example : he did complicated pen-and-ink drawings

Word : Study
Example : he knocked on the door of the study

Word : Study
Example : he made several studies before the final painting

w_1, w_2	c_1, c_2	IsHypernym(w_1, w_2) ?
Study, Room	He knocked on the door of the **study** The **rooms** were small but comfortable	✓
Study, Drawing	He made 2 **studies** before the painting He did pen-and-ink **drawings**	✓

WHiC : Negative Examples

Word : Room
Example : the rooms were small but comfortable

Word : Drawing
Example : he did complicated pen-and-ink drawings

Word : Study
Example : he knocked on the door of the study

Word : Study
Example : he made several studies before the final painting

w_1, w_2	c_1, c_2	IsHypernym(w_1, w_2) ?
Study, Room	He knocked on the door of the **study** The **rooms** were small but comfortable	✓
Study, Drawing	He made 2 **studies** before the painting He did pen-and-ink **drawings**	✓
Study, Room	He knocked on the door of the **study** He did pen-and-ink **drawings**	✗

WHiC : Negative Examples

Word : Room
Example : the rooms were small but comfortable

Word : Study
Example : he knocked on the door of the study

Word : Drawing
Example : he did complicated pen-and-ink drawings

Word : Study
Example : he made several studies before the final painting

w_1, w_2	c_1, c_2	IsHypernym(w_1, w_2) ?
Study, Room	He knocked on the door of the **study** The **rooms** were small but comfortable	✓
Study, Drawing	He made 2 **studies** before the painting He did pen-and-ink **drawings**	✓
Study, Room	He knocked on the door of the **study** He did pen-and-ink **drawings**	✗

WHiC : Negative Examples

Word : Room
Example : the rooms were small but comfortable

Word : Study
Example : he knocked on the door of the study

Word : Drawing
Example : he did complicated pen-and-ink drawings

Word : Study
Example : he made several studies before the final painting

w_1, w_2	c_1, c_2	IsHypernym(w_1, w_2) ?
Study, Room	He knocked on the door of the **study** The **rooms** were small but comfortable	✓
Study, Drawing	He made 2 **studies** before the painting He did pen-and-ink **drawings**	✓
Study, Room	He knocked on the door of the **study** He did pen-and-ink **drawings**	✗
Room, Study	The **rooms** were small but comfortable He knocked on the door of the **study**	✗

Dataset Desiderata

- Test sensitivity of models to changing contexts ✓
- Differentiate between **hypernymy** and **semantic similarity** ✓
- Large enough to train supervised learning models on ✓
  - Seed vocab = 10k most frequent words from Wikipedia
  - 22000 examples distributed over 6000 word pairs

Outline

- WHiC : A Dataset for Hypernymy Detection in Context
  - Dataset Desiderata
  - Dataset Creation
- Two Models of Context Representation
- Experiments and Analysis

Outline

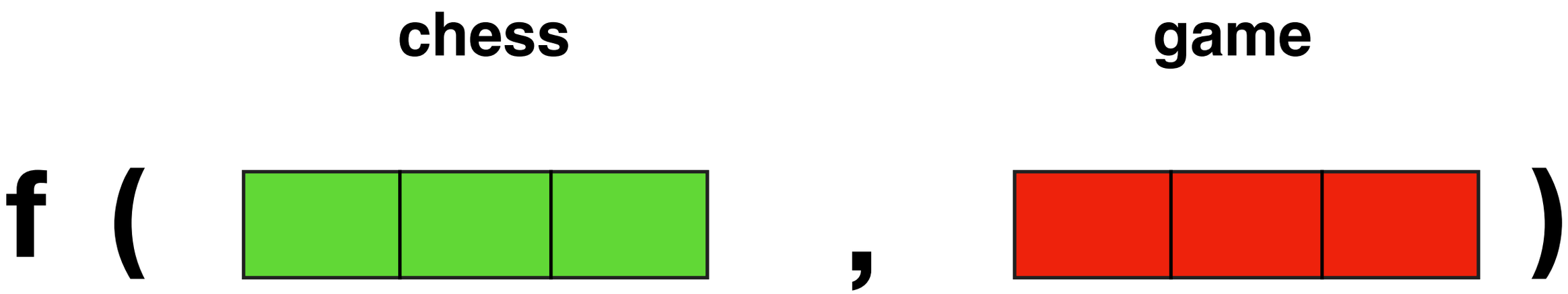
- WHiC : A Dataset for Hypernymy Detection in Context
  - Dataset Desiderata
  - Dataset Creation
- Two Models of Context Representation
- Experiments and Analysis

Supervised Hypernymy Detection

chess **game**

Supervised Hypernymy Detection

Supervised Hypernymy Detection



Baroni et al., 2012; Roller et al., 2014; Weeds et al., 2014 *inter alia*

Supervised Hypernymy Detection

chess **game**

f (,)

x . y **y - x** [**x ; y**]

Baroni et al., 2012; Roller et al., 2014; Weeds et al., 2014 *inter alia*

Hypernymy-Feature Detector

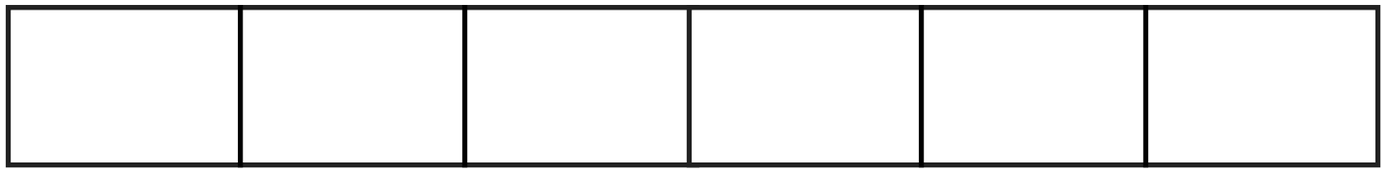
Roller and Erk; EMNLP 2016

Game

Word (w)

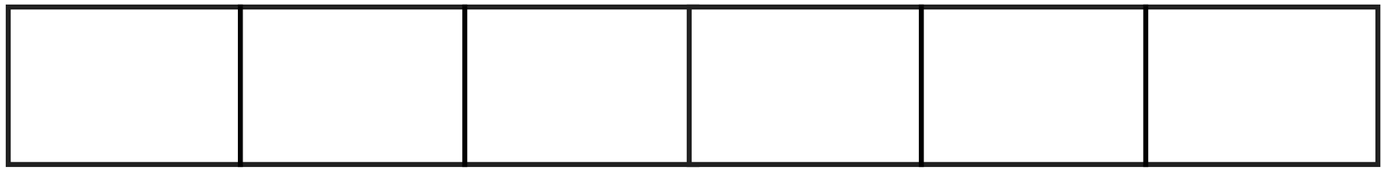
Game

Word (w)



The hunters hunted the **game**

Contextualized Word (w_c)



Game

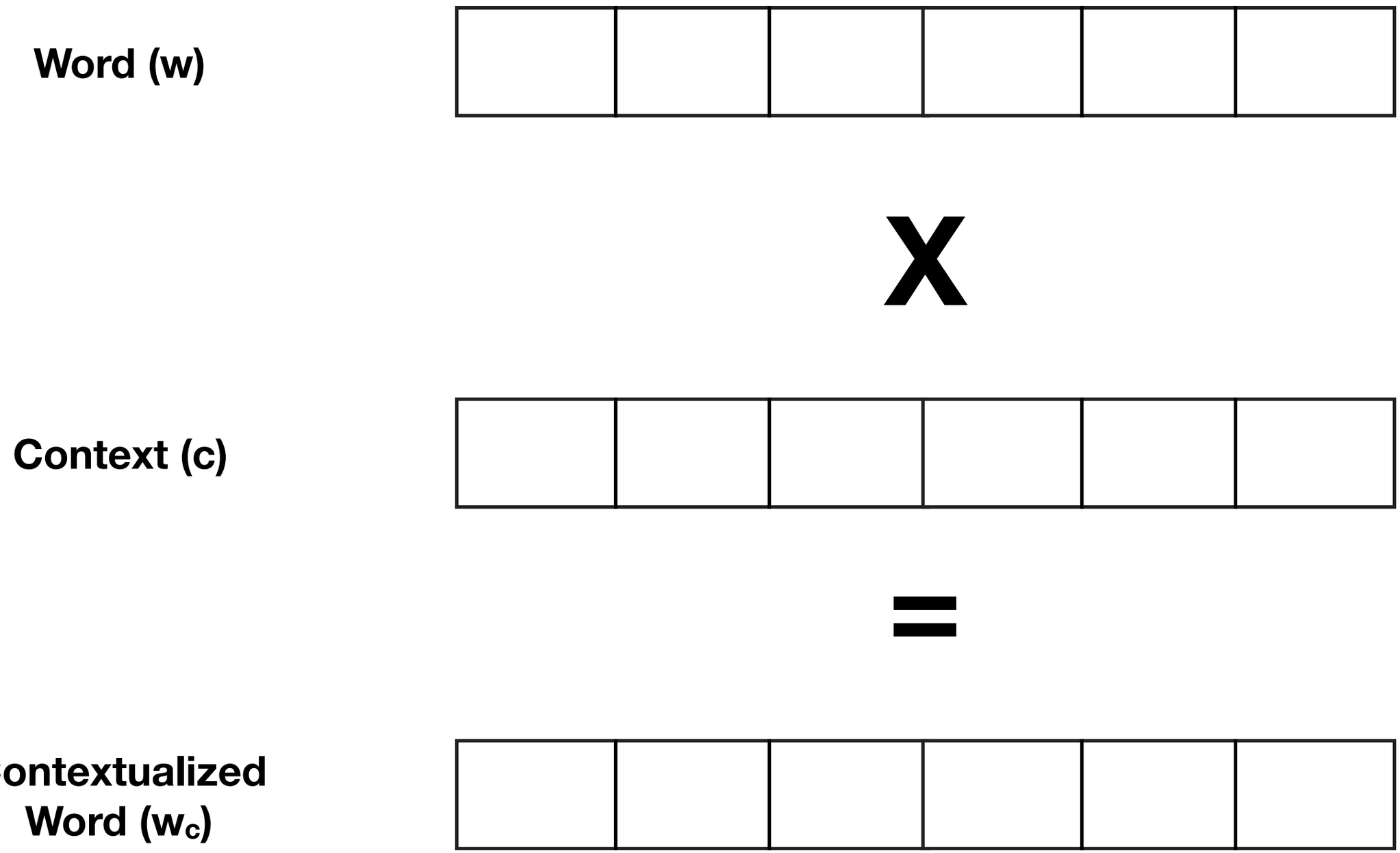
Word (w)

The hunters hunted the game

Context (c)

The hunters hunted the **game**

Contextualized Word (w_c)



Erk and Padó, EMNLP 2008; Thater et al., IJCNLP 2011; Dinu et al., NAACL 2012

Game

Word (w)

The hunters hunted the game

Context (c)

The hunters hunted the **game**

Contextualized Word (w_c)

Representing Context

Two complementary approaches

Representing Context

Two complementary approaches

- **Pooling :** Combining word type representations into context representations
  - Mean, min, and max

Representing Context : Pooling

Representing Context : Pooling

the	0.3	0.6	-0.1
river	1.5	-2.5	0
bank	-1	0.2	1.8

Representing Context : Pooling

the	0.3	0.6	-0.1
river	1.5	-2.5	0
bank	-1	0.2	1.8

$\vec{c}_{l,mean}$

0.27	-0.57	0.57

Representing Context : Pooling

the	0.3	0.6	-0.1
river	1.5	-2.5	0
bank	-1	0.2	1.8

$\vec{c}_{l,mean}$	0.27	-0.57	0.57
$\vec{c}_{l,min}$	-1	-2.5	-0.1

Representing Context : Pooling

the	0.3	0.6	-0.1
river	1.5	-2.5	0
bank	-1	0.2	1.8

$\vec{c}_{l,mean}$	0.27	-0.57	0.57
$\vec{c}_{l,min}$	-1	-2.5	-0.1
$\vec{c}_{l,max}$	1.5	0.6	1.8

Representing Context : Pooling

the	0.3	0.6	-0.1
river	1.5	-2.5	0
bank	-1	0.2	1.8

$\vec{c}_{l,mean}$	0.27	-0.57	0.57
$\vec{c}_{l,min}$	-1	-2.5	-0.1
$\vec{c}_{l,max}$	1.5	0.6	1.8

Tang et al., ACL 2014; Hovy, ACL 2015

Representing Context : Pooling

the	0.3	0.6	-0.1
river	1.5	-2.5	0
bank	-1	0.2	1.8

$\vec{c}_{l,mean}$	0.27	-0.57	0.57
$\vec{c}_{l,min}$	-1	-2.5	-0.1
$\vec{c}_{l,max}$	1.5	0.6	1.8

$$\vec{c}_l = [\vec{c}_{l,mean} ; \vec{c}_{l,min} ; \vec{c}_{l,max}]$$

Tang et al., ACL 2014; Hovy, ACL 2015

Representing Context

Two complementary approaches

- **Pooling :** Combining word type representations into context representations
  - Mean, min, and max

Representing Context

Two complementary approaches

- **Pooling :** Combining word type representations into context representations
  - Mean, min, and max
- **RNN Composition** : Embed words and contexts jointly

Representing Context : Context2Vec

Melamud et al; CoNLL 2016

Representing Context : Context2Vec

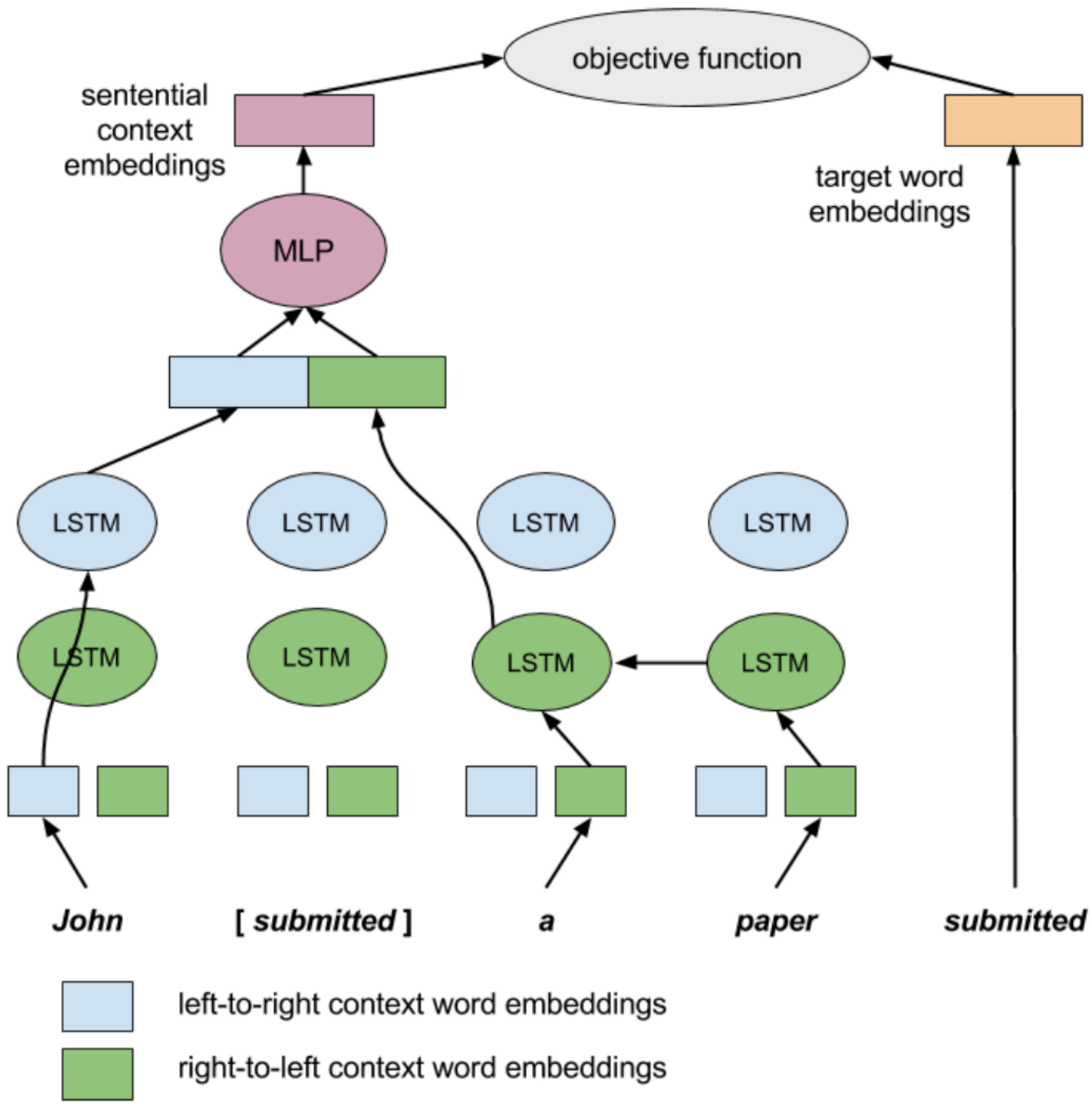
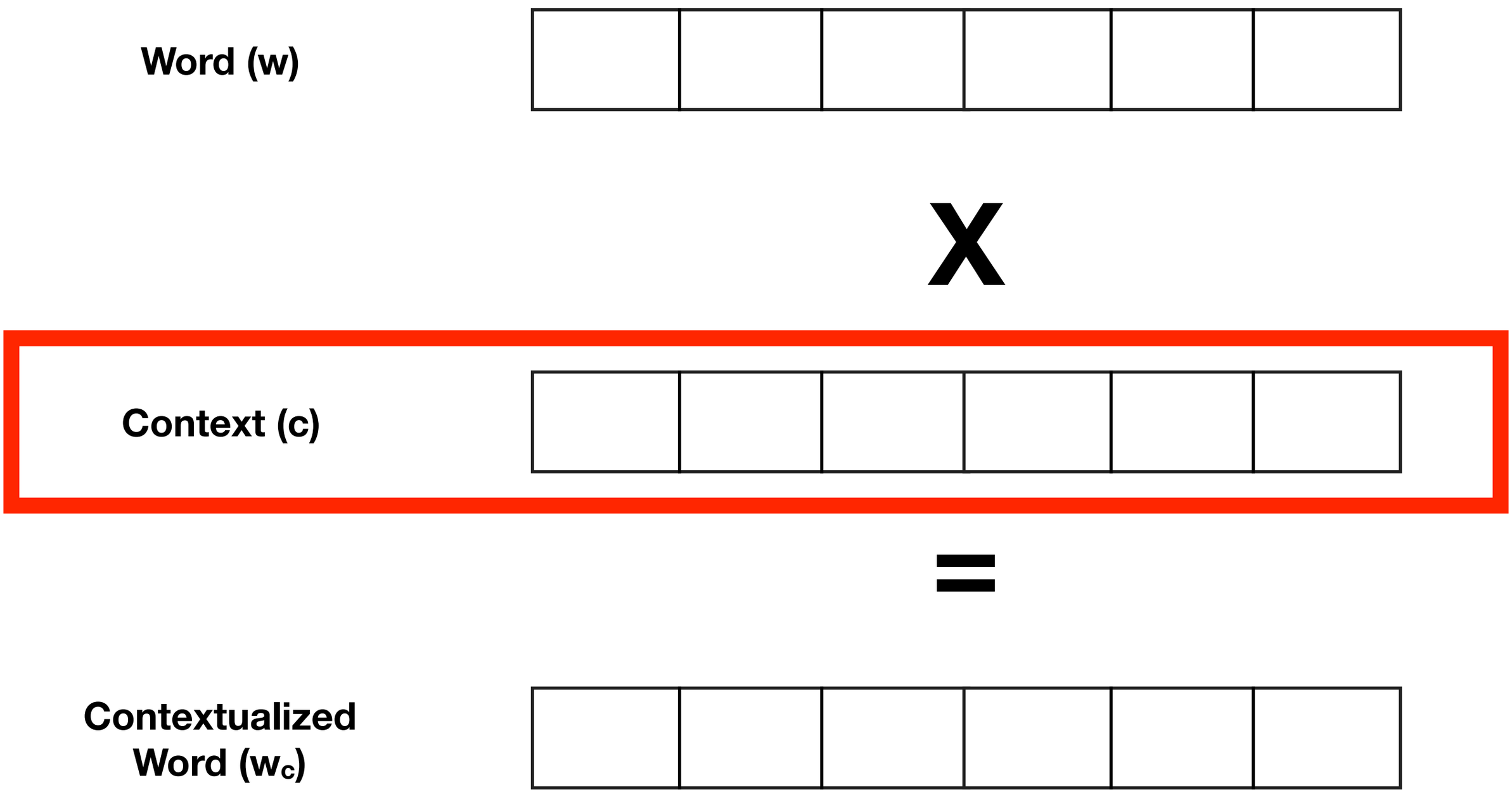
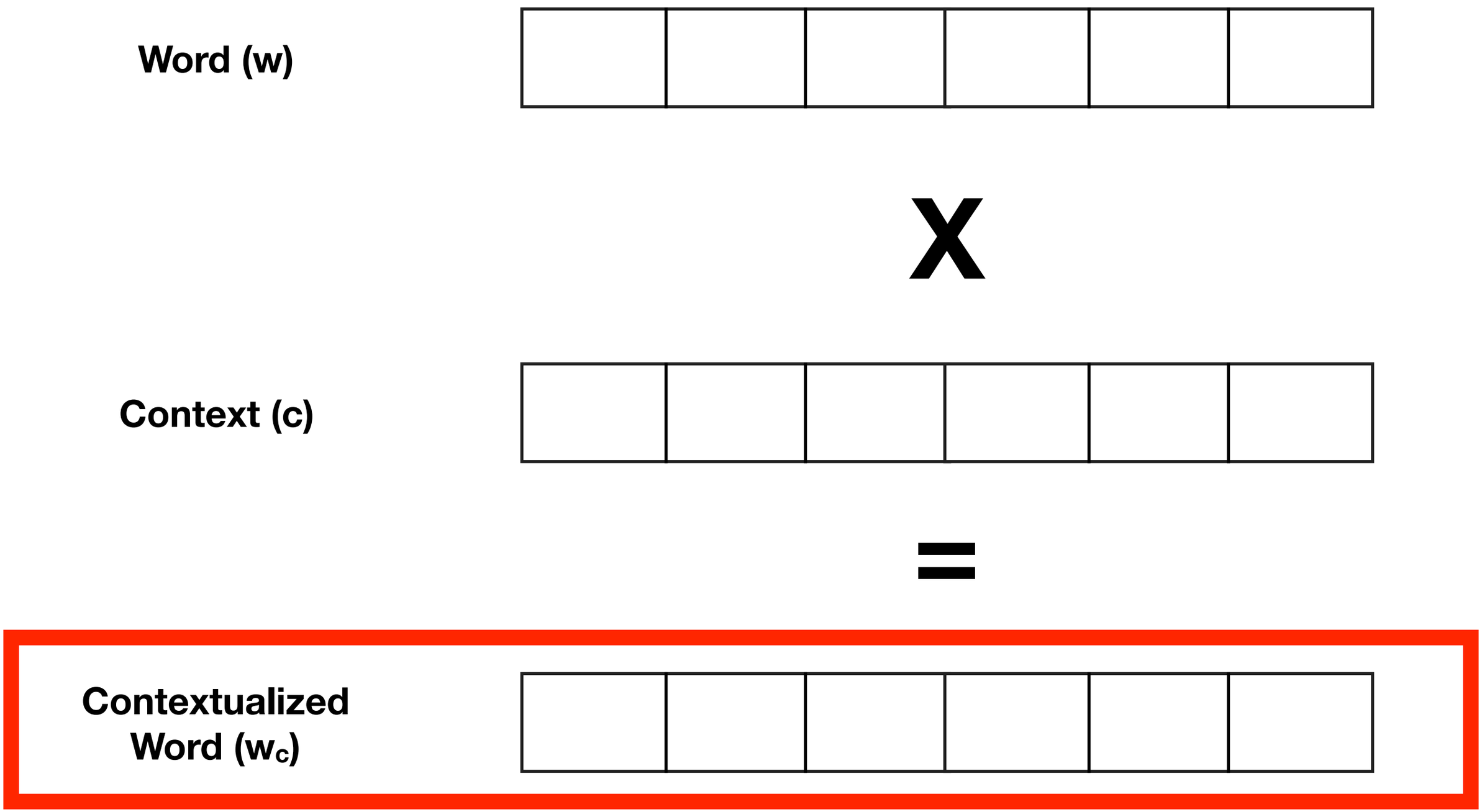


Image courtesy : context2vec: Learning Generic Context Embedding with Bidirectional LSTM, Melamud et al., CoNLL 2016


Word (w)
X
Context (c)
=
Contextualized Word (w_c)


Word (w)
X
Context (c)
=
Contextualized Word (w_c)

Putting it all together ..

- **Task** : Context Aware Hypernymy Detection

Putting it all together ..

- **Task** : Context Aware Hypernymy Detection
- **Dataset** : WHiC

Putting it all together ..

- **Task** : Context Aware Hypernymy Detection
- **Dataset** : WHiC
- **Classifier** : H-Feature Detector

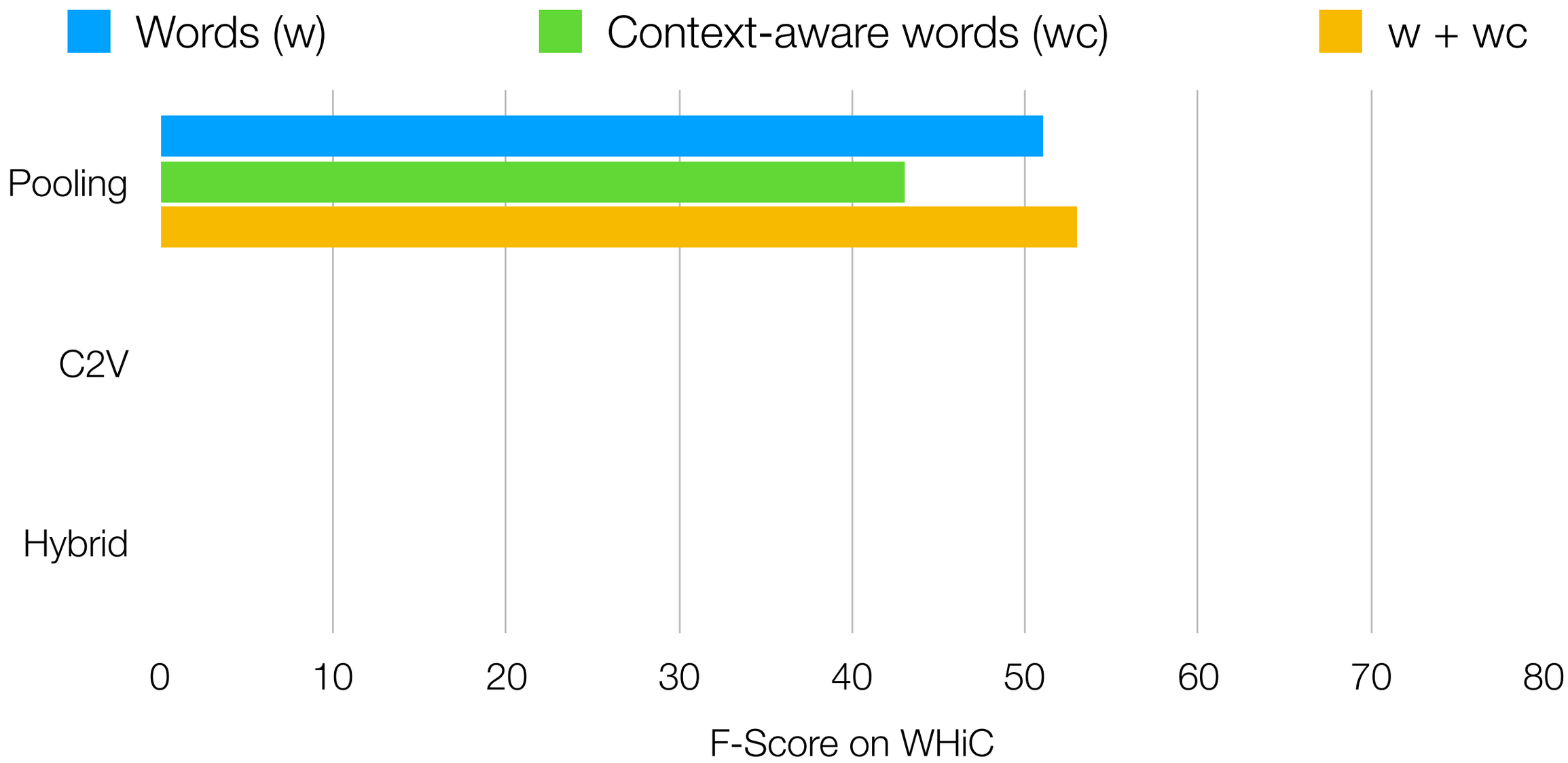
Putting it all together ..

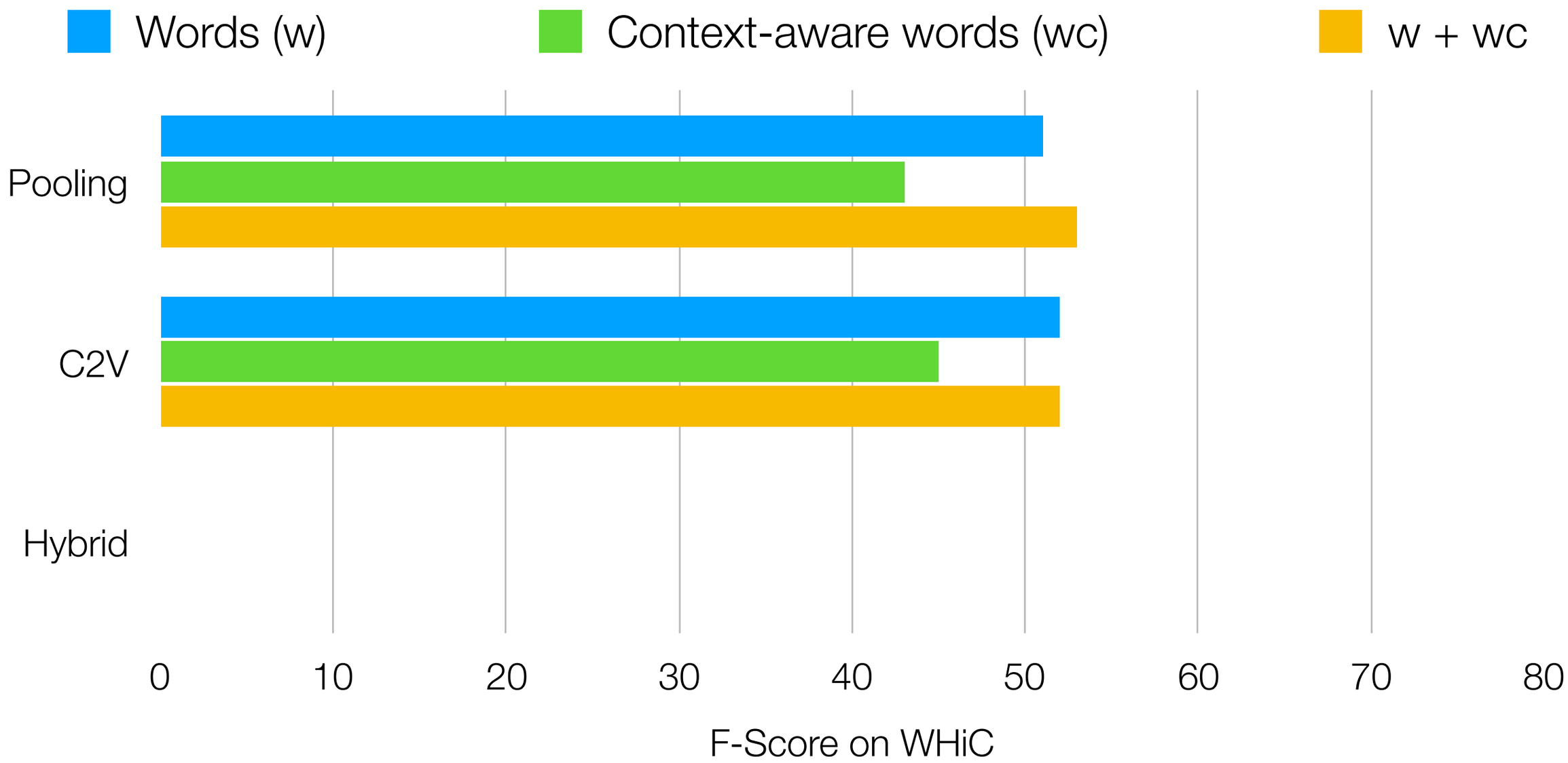
- **Task** : Context Aware Hypernymy Detection
- **Dataset** : WHiC
- **Classifier** : H-Feature Detector
- **Input to classifier** : Context aware word representations
  - Pooling
  - Context2Vec

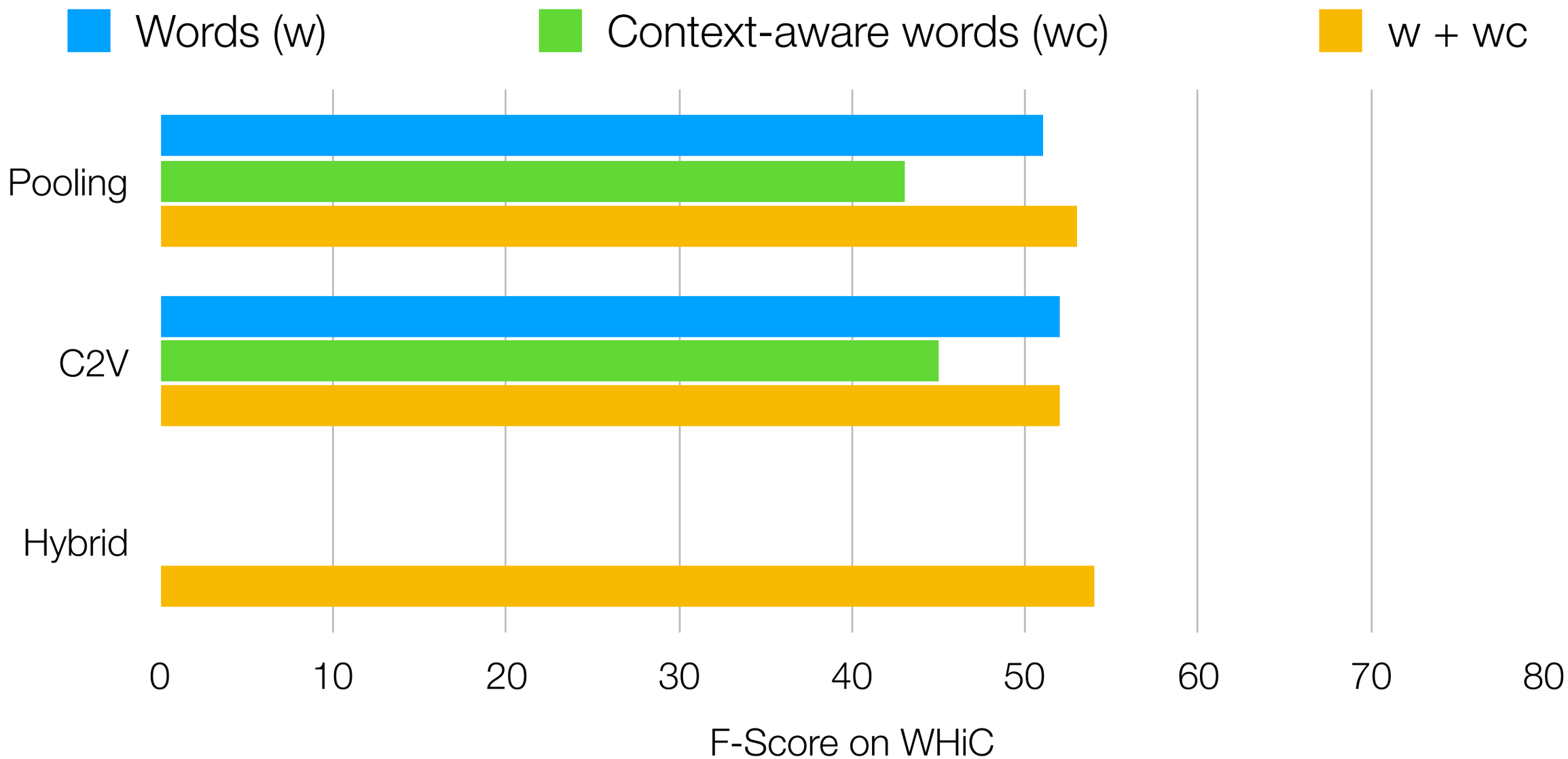
Outline

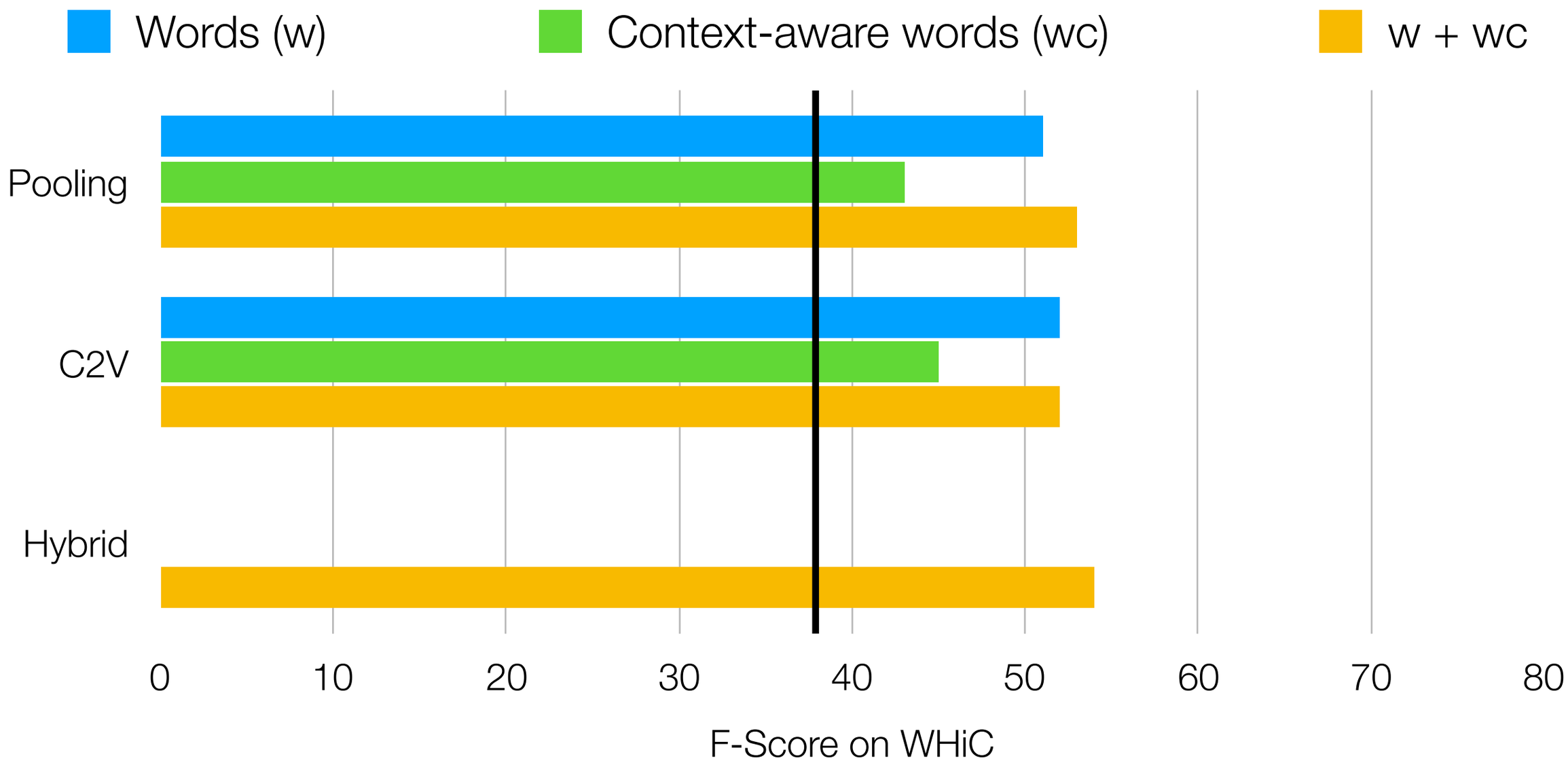
- WHiC : A Dataset for Hypernymy Detection in Context
  - Dataset Desiderata
  - Dataset Creation
- Two Models of Context Representation
- Experiments and Analysis

Overall Results

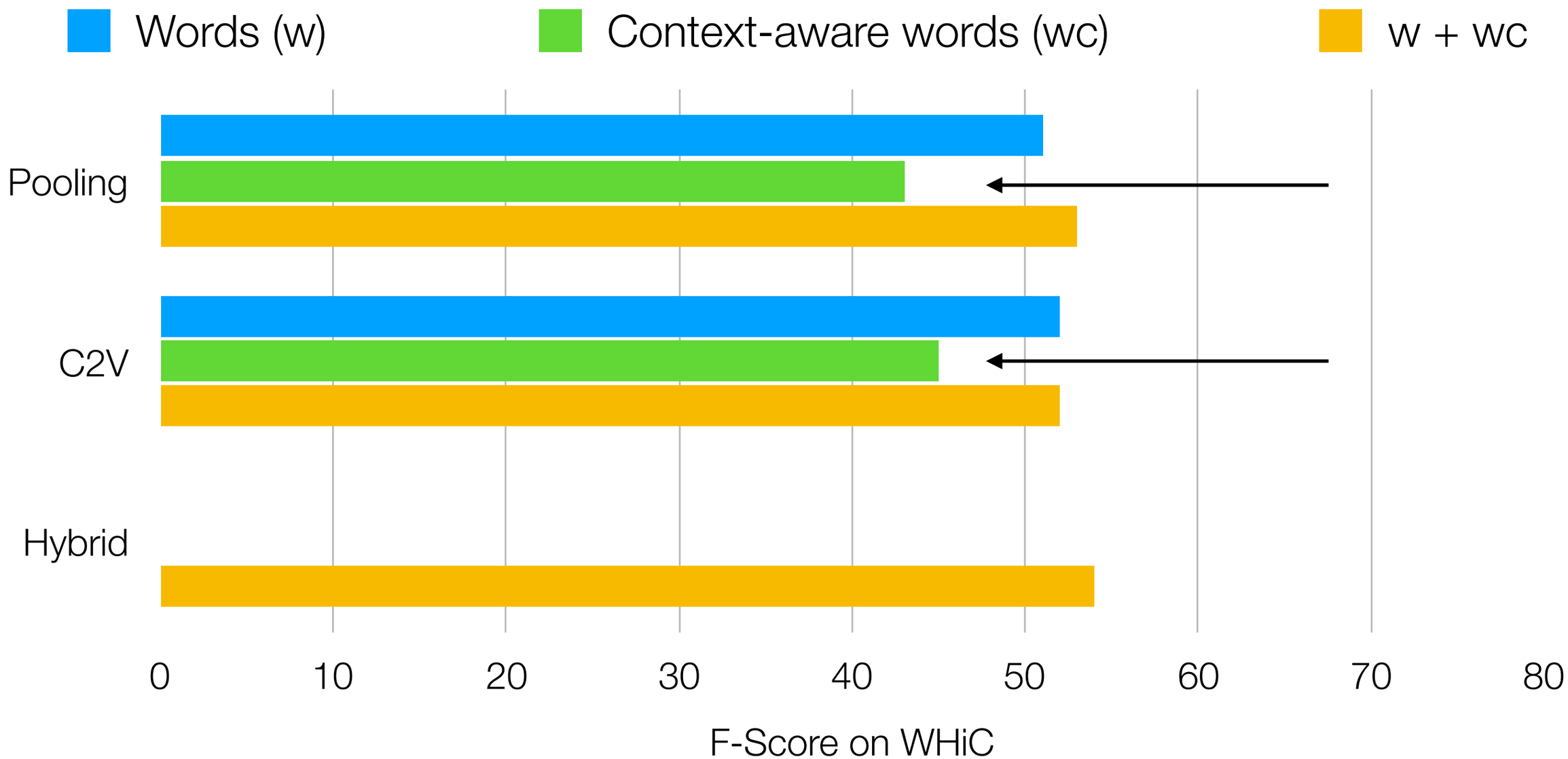

Words (w)
Context-aware words (wc)
w + wc
Pooling
C2V
Hybrid
0
10
20
30
40
50
60
70
80
F-Score on WHiC


Words (w)
Context-aware words (wc)
w + wc
Pooling
C2V
Hybrid
0
10
20
30
40
50
60
70
80
F-Score on WHiC

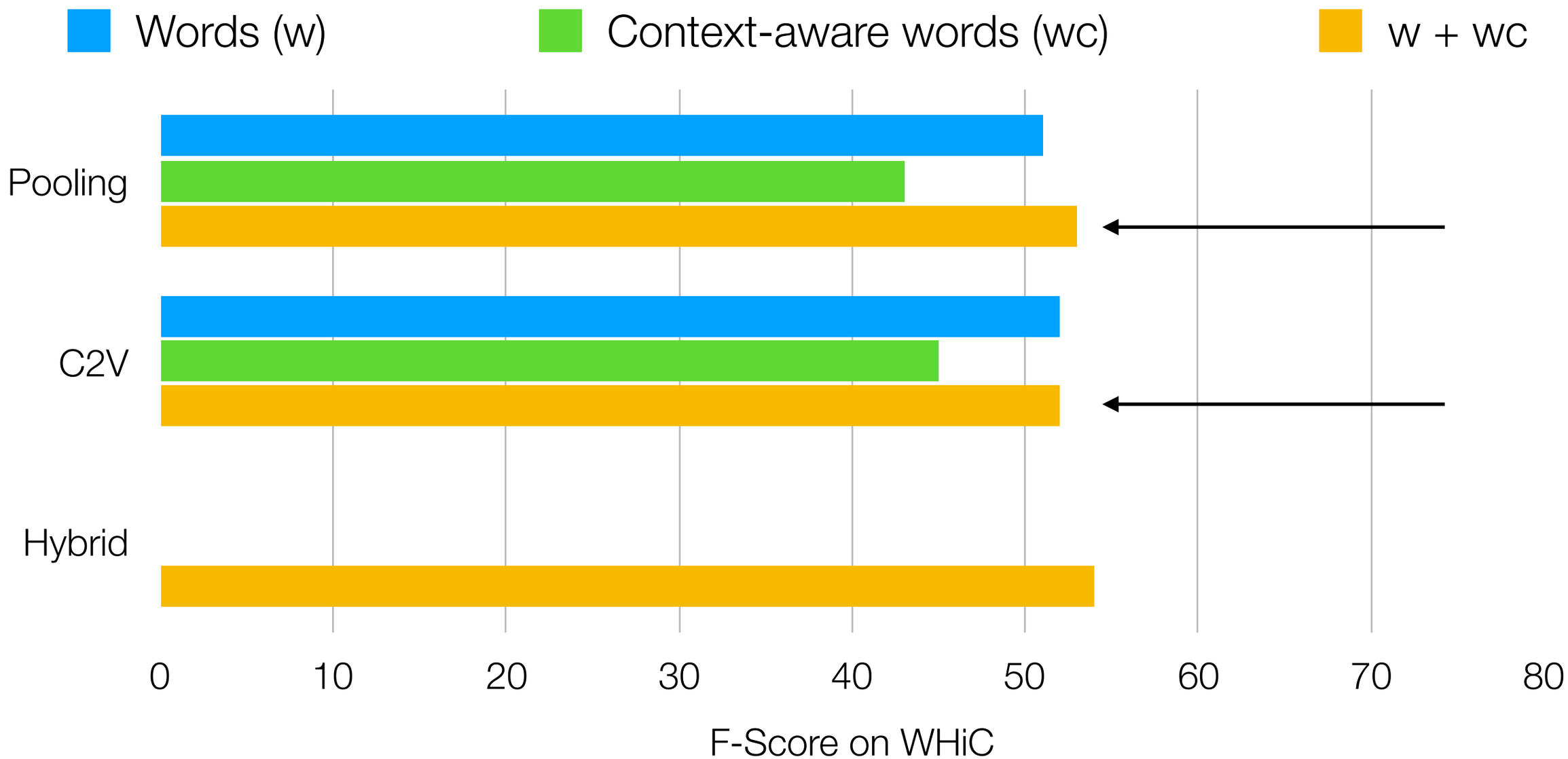

Words (w)
Context-aware words (wc)
w + wc
Pooling
C2V
Hybrid
0
10
20
30
40
50
60
70
80
F-Score on WHiC



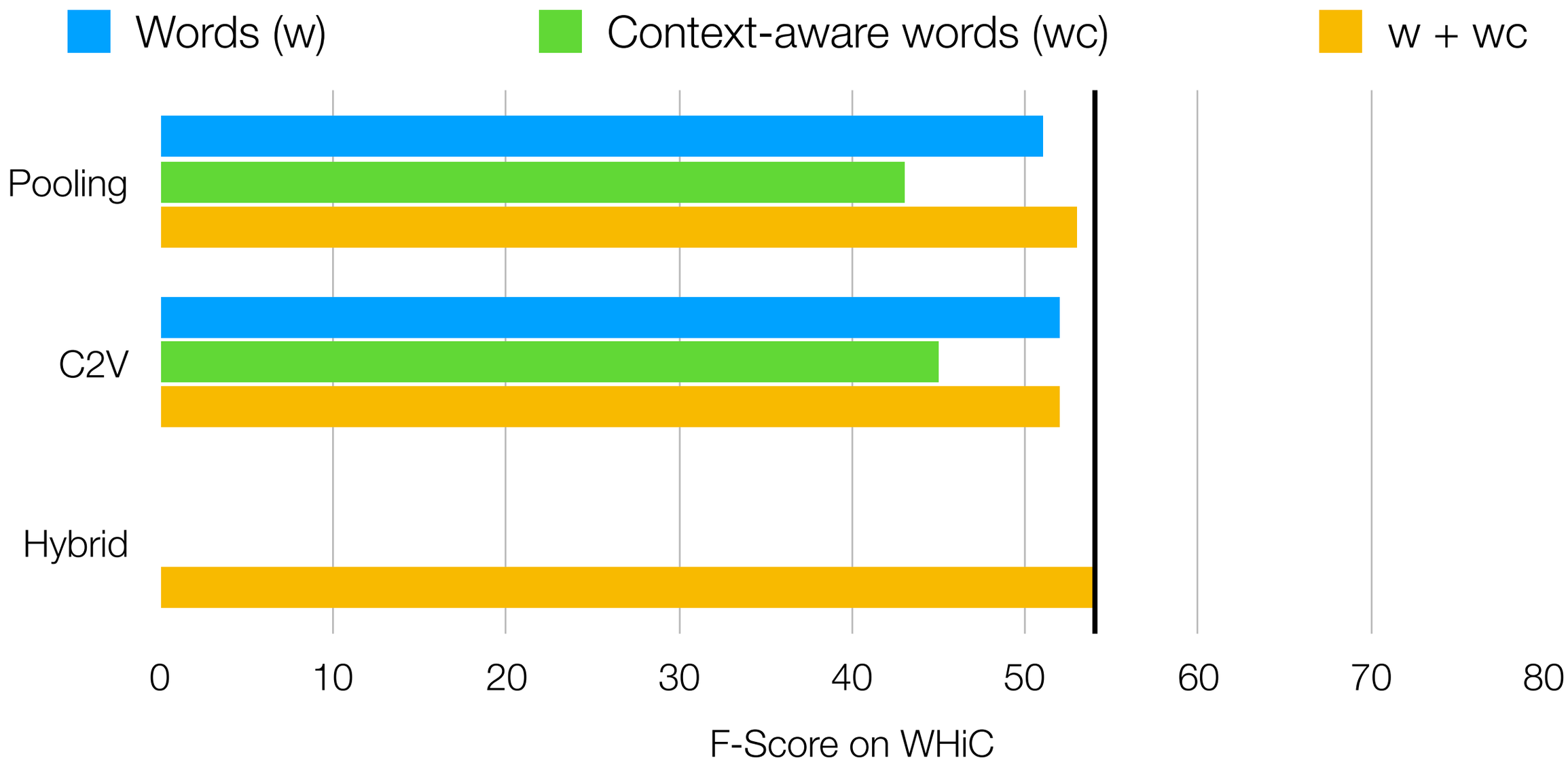
- All models > “All True” Baseline



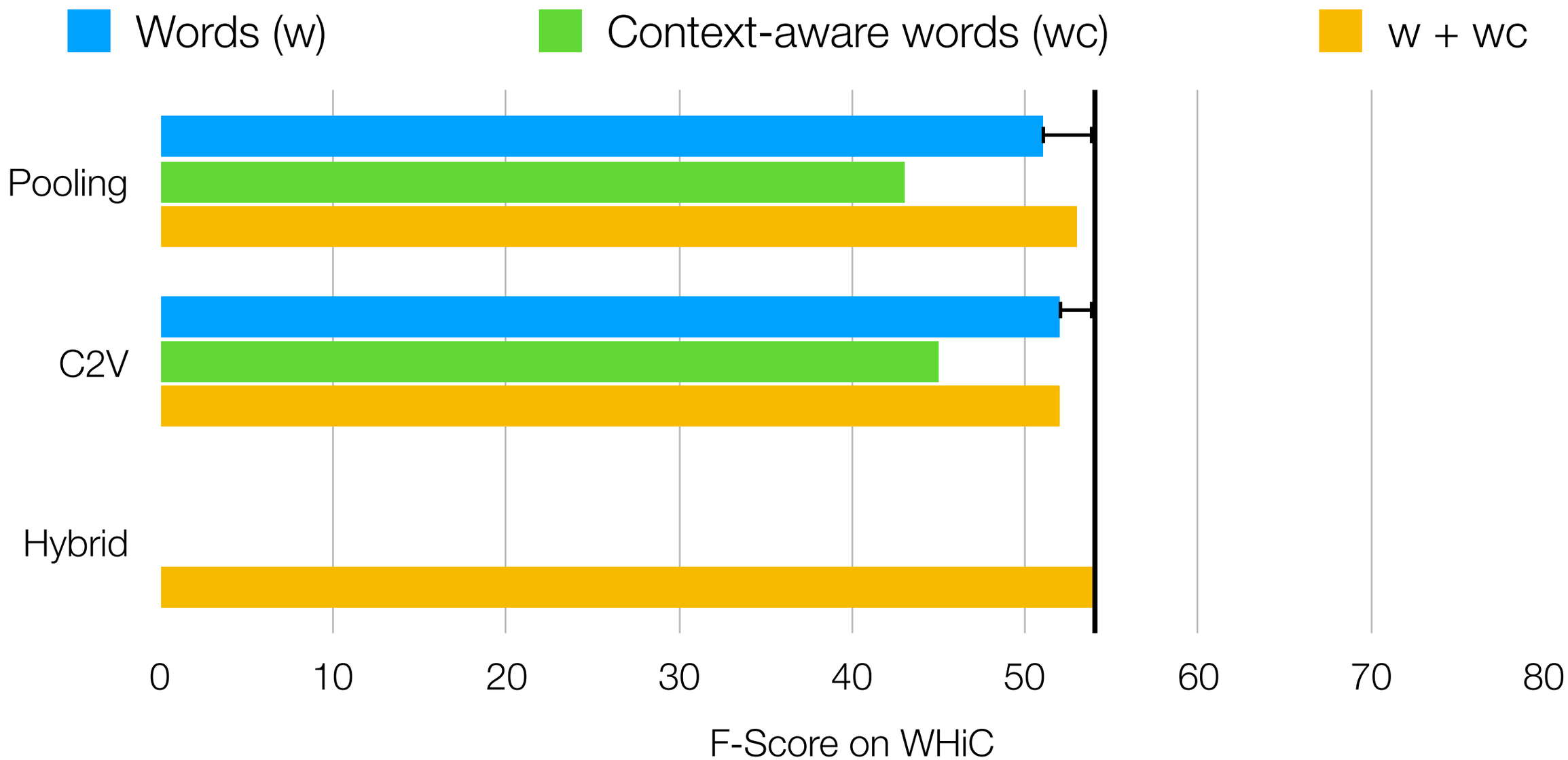
- All models > “All True” Baseline
- Context aware representations help ..



- All models > “All True” Baseline
- Context aware representations help ..



- All models > “All True” Baseline
- Context aware representations help ..
- Hybrid model works best



- All models > “All True” Baseline
- Context aware representations help ..
- Hybrid model works best
- Small improvements over context agnostic models

Analysis : Sensitivity to Context

Analysis : Sensitivity to Context

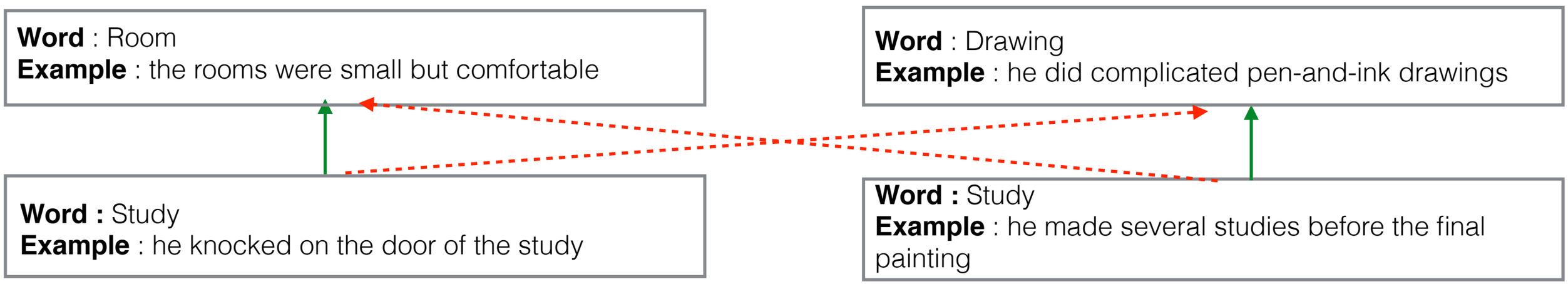
Word : Room
Example : the rooms were small but comfortable

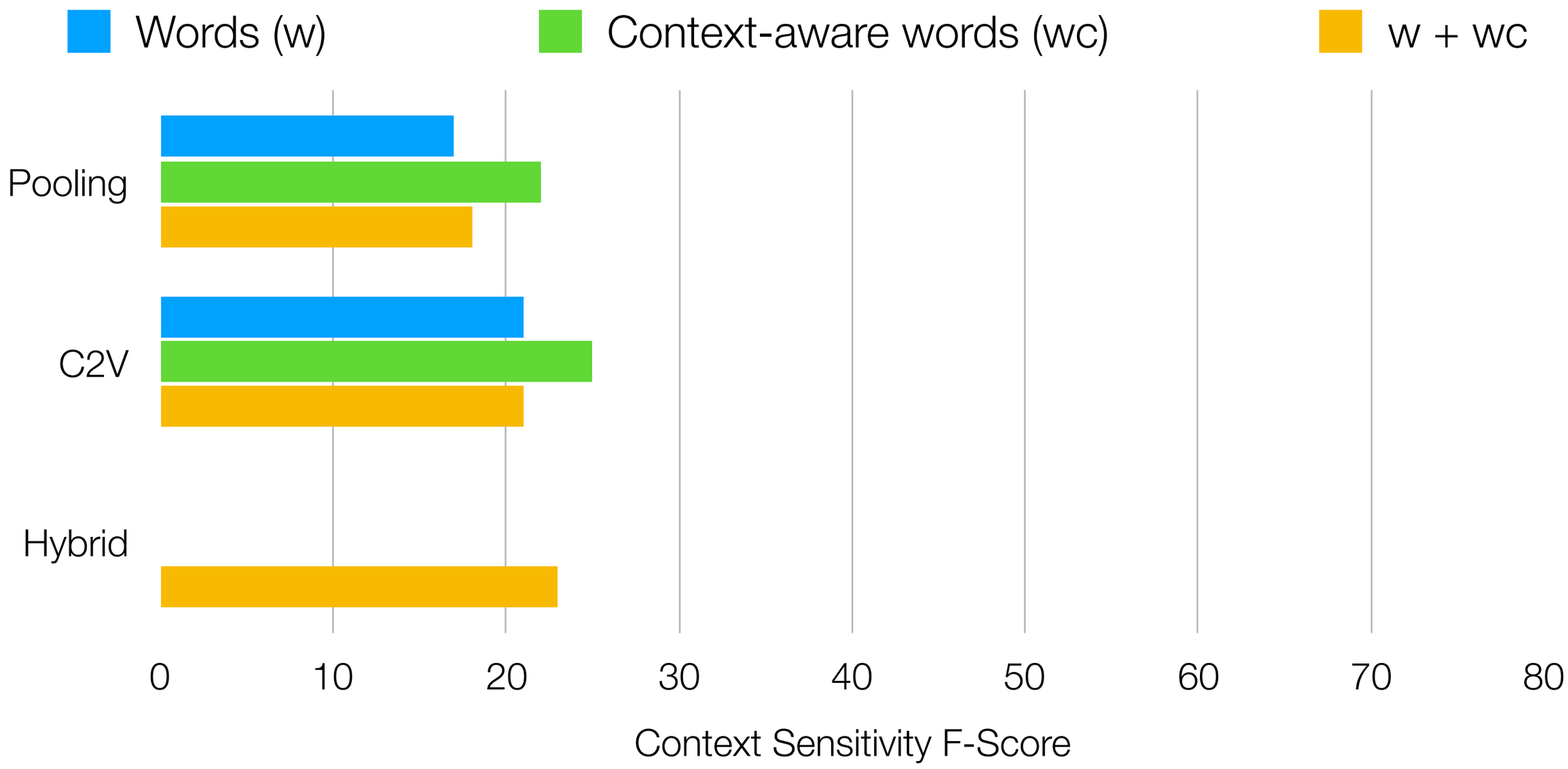
Word : Study
Example : he knocked on the door of the study

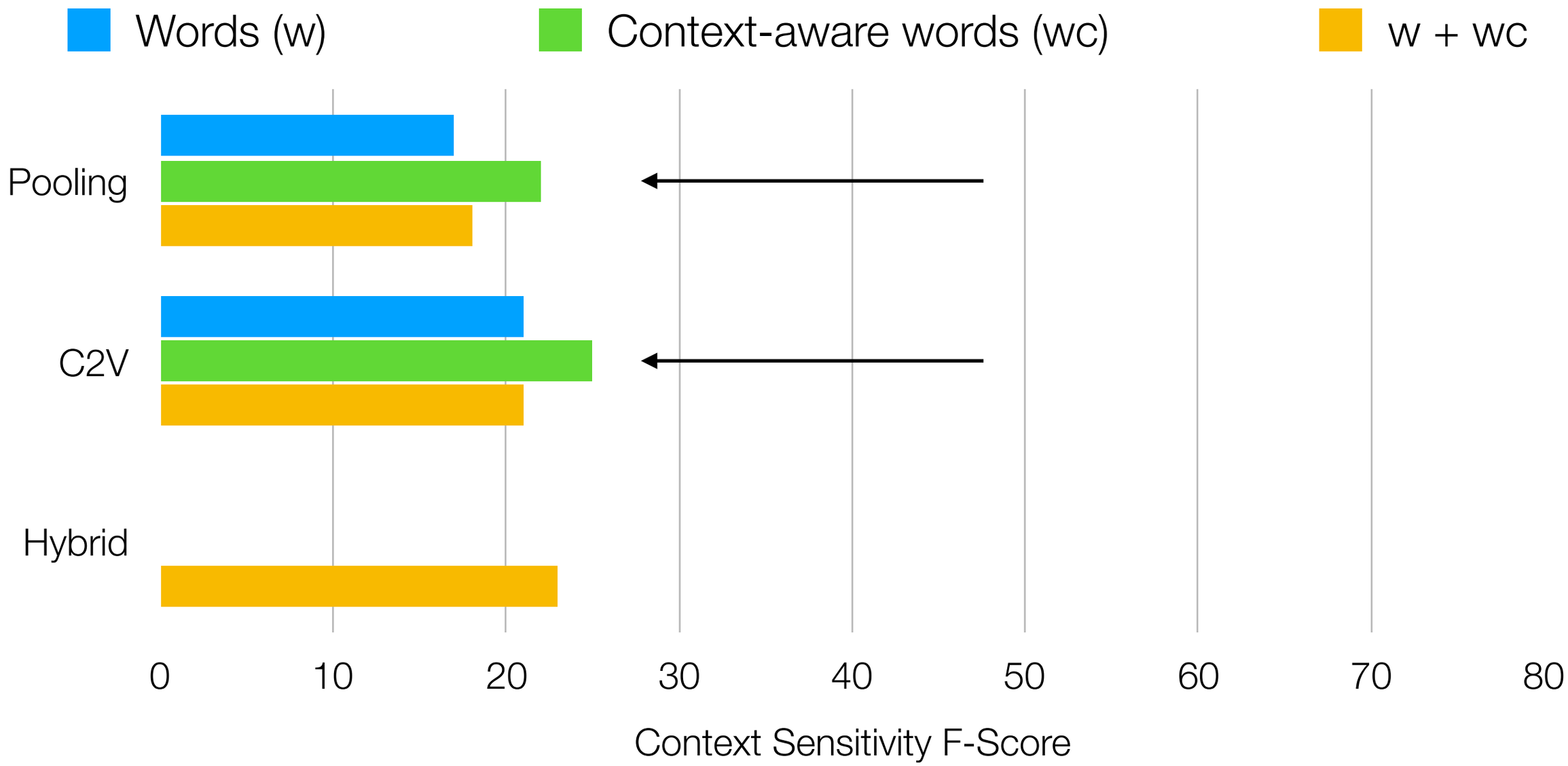
Word : Drawing
Example : he did complicated pen-and-ink drawings

Word : Study
Example : he made several studies before the final painting

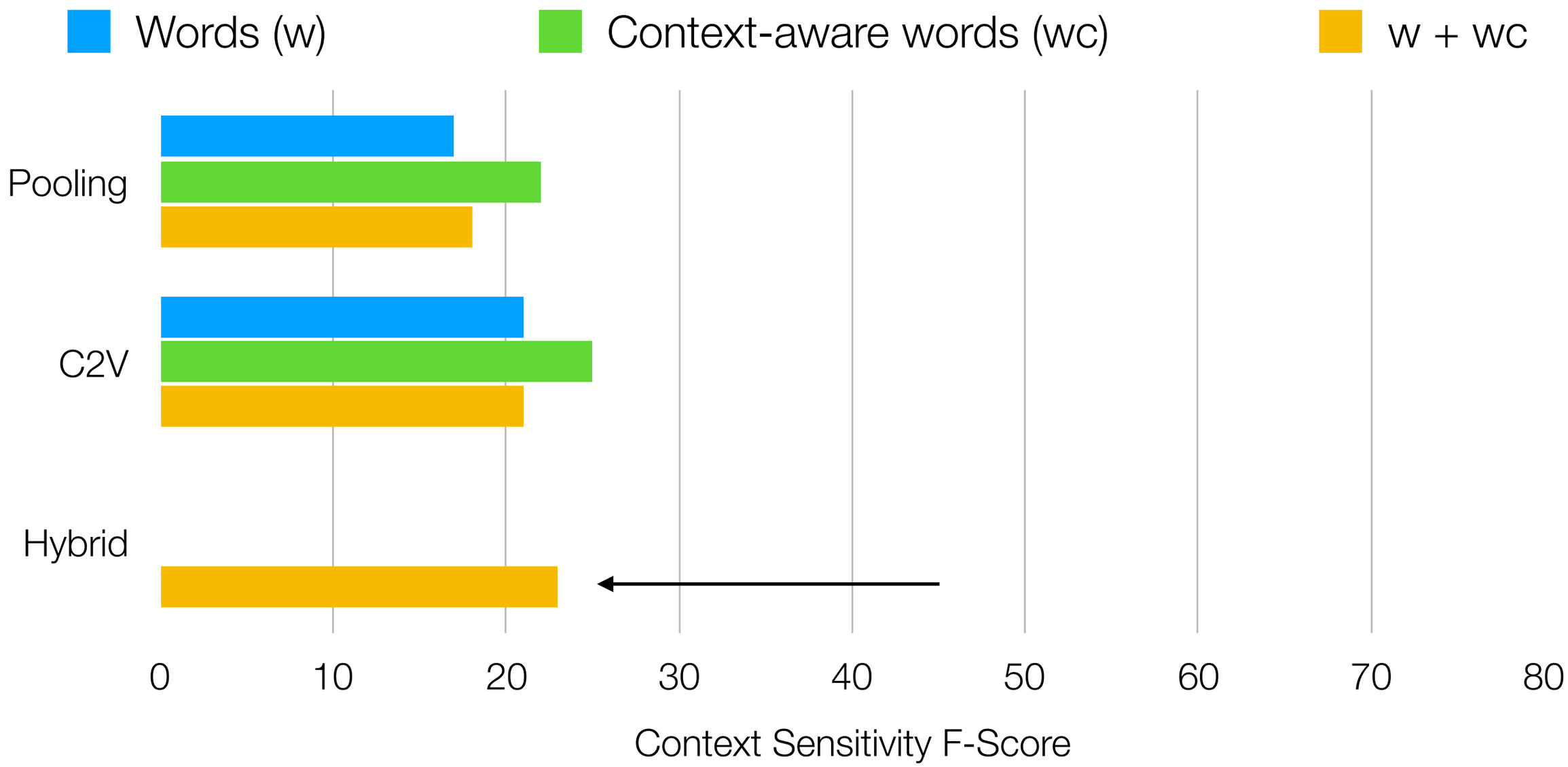
Analysis : Sensitivity to Context




Words (w)
Context-aware words (wc)
w + wc
Pooling
C2V
Hybrid
0
10
20
30
40
50
60
70
80
Context Sensitivity F-Score



- Context aware representations do better on this subset



- Context aware representations do better on this subset ..
- .. followed by the hybrid model

Analysis : Directionality

Analysis : Directionality

Word : Room
Example : the rooms were small but comfortable

Word : Study
Example : he knocked on the door of the study

Word : Drawing
Example : he did complicated pen-and-ink drawings

Word : Study
Example : he made several studies before the final painting

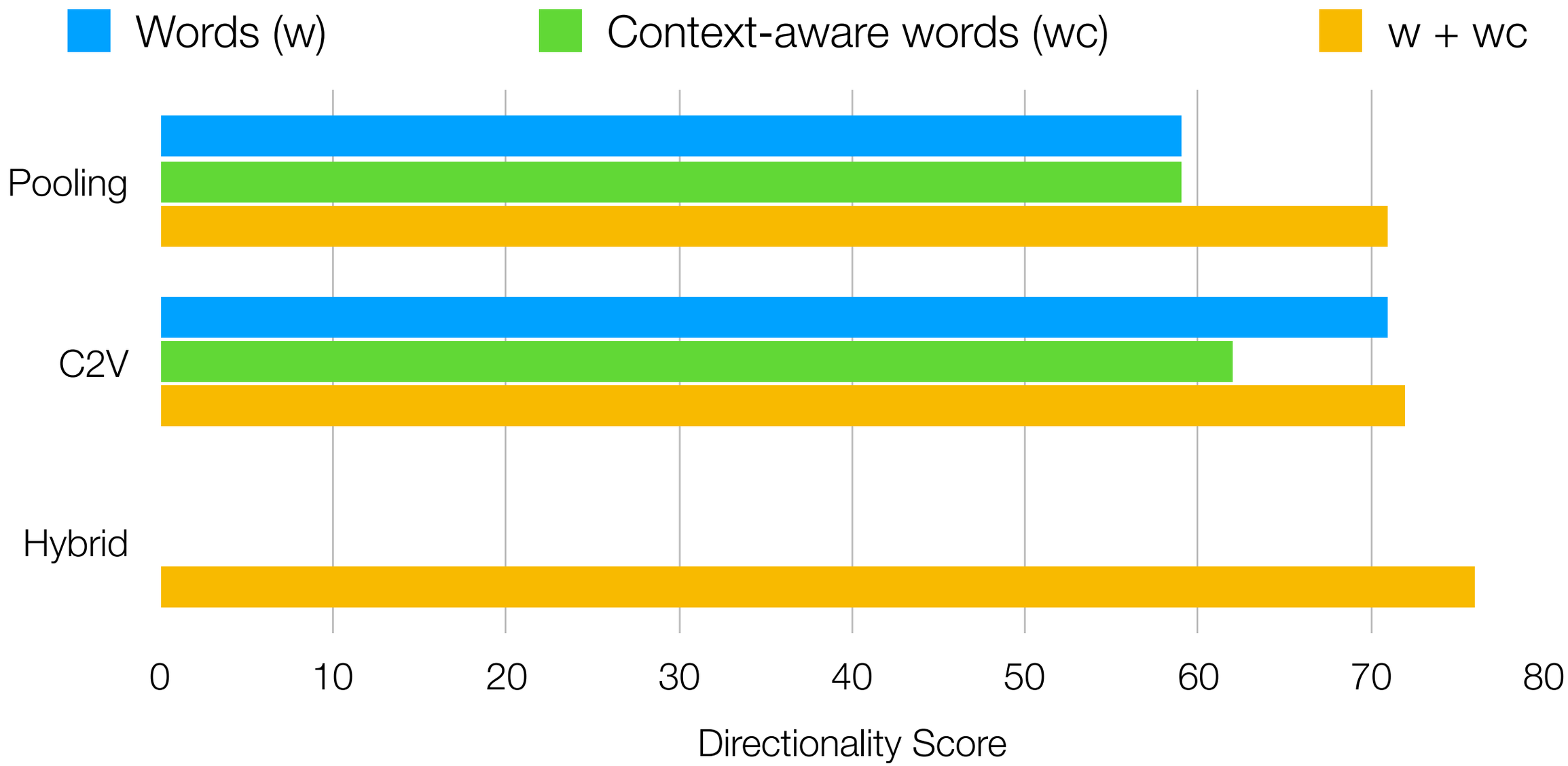
Analysis : Directionality

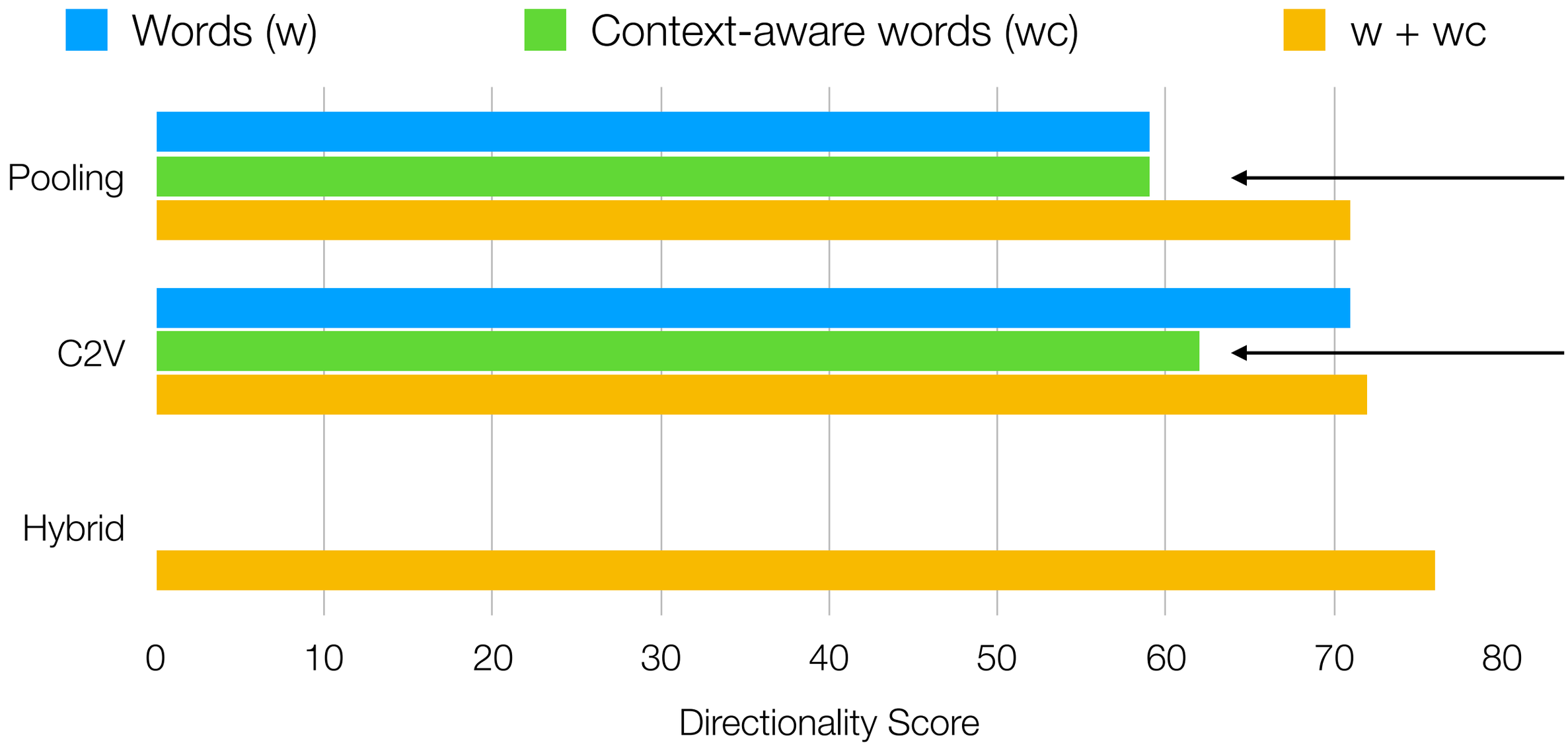
Word : Room
Example : the rooms were small but comfortable

Word : Study
Example : he knocked on the door of the study

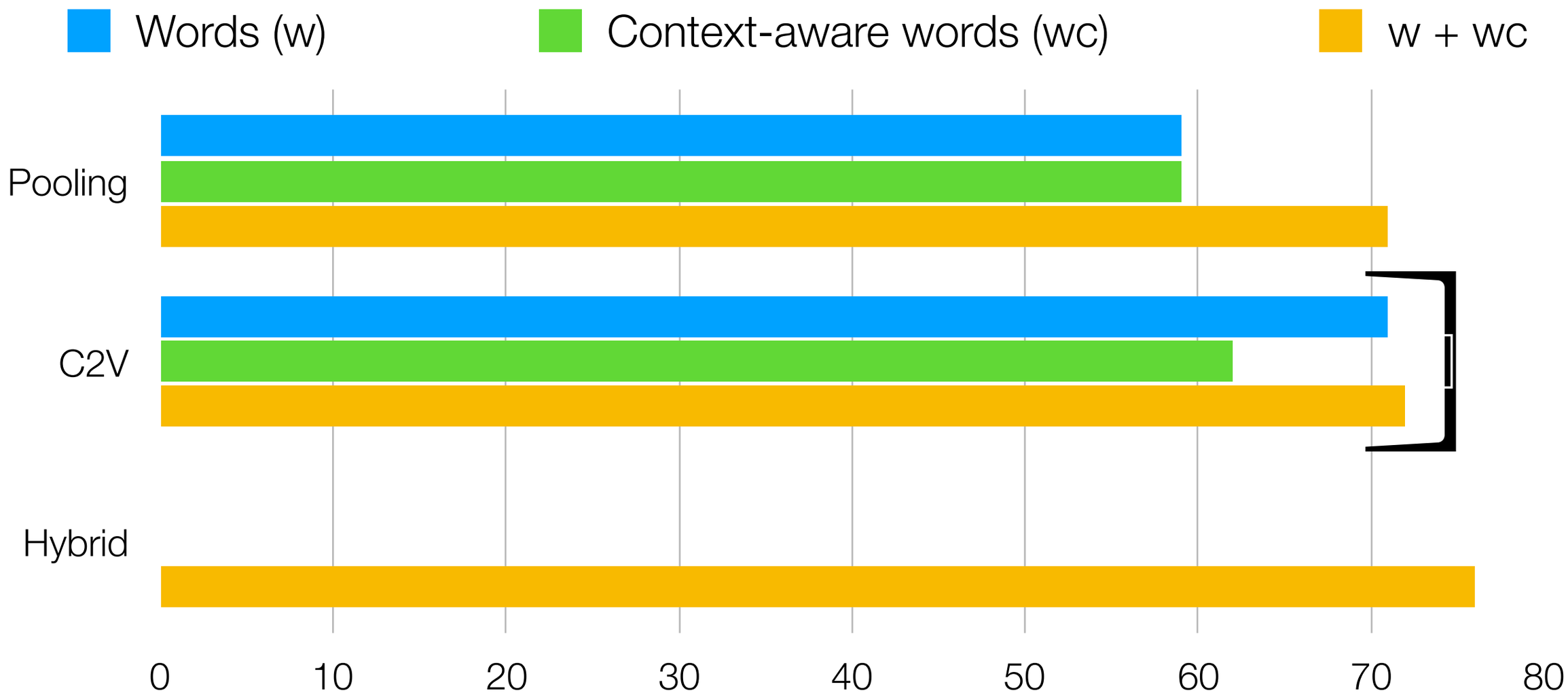
Word : Drawing
Example : he did complicated pen-and-ink drawings

Word : Study
Example : he made several studies before the final painting

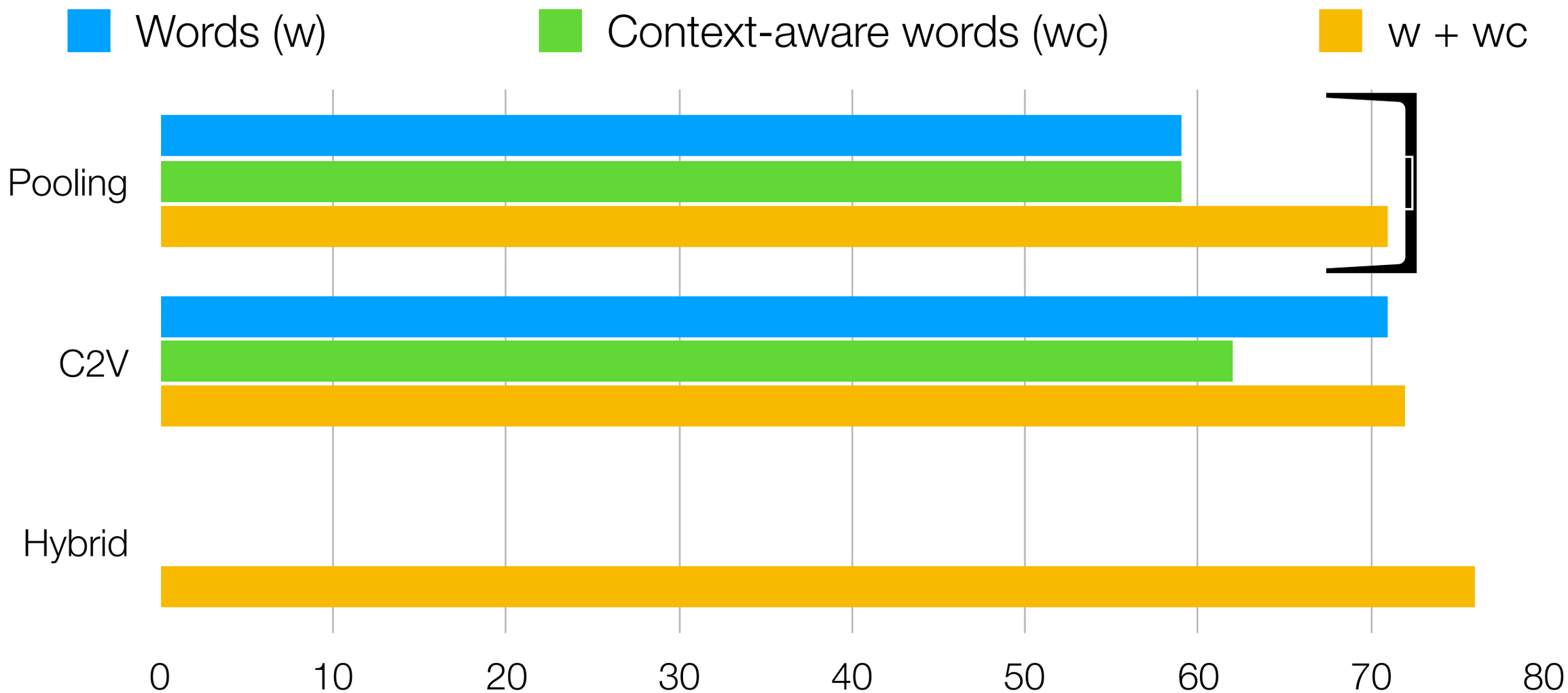

Words (w)
Context-aware words (wc)
w + wc
Pooling
C2V
Hybrid
0
10
20
30
40
50
60
70
80
Directionality Score



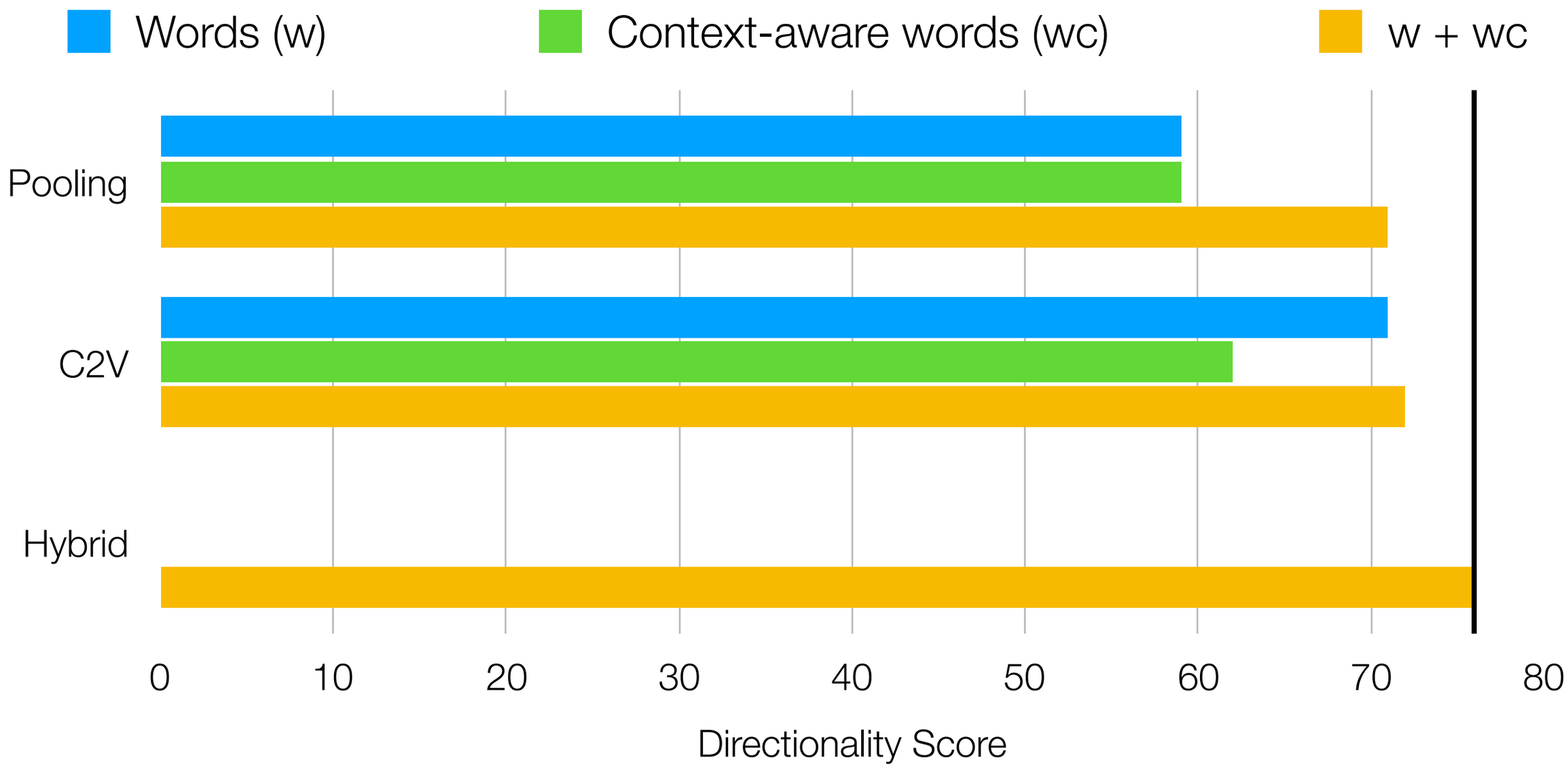
- Context-aware representations do worse



- Context-aware representations do worse
- Context2Vec models are better than Pooling models



- Context-aware representations do worse
- Context2Vec models are better than Pooling models



- Context-aware representations do worse
- Context2Vec models are better than Pooling models
- Hybrid model is best - mirroring performance on the full dataset

Results Summary

- Context aware hypernymy detection is **challenging**!

Results Summary

- Context aware hypernymy detection is **challenging**!
- Context aware representations **help** in capturing **context**
  - Do **worse** at capturing **asymmetry**

Results Summary

- Context aware hypernymy detection is **challenging**!
- Context aware representations **help** in capturing **context**
  - Do **worse** at capturing **asymmetry**
- **Hybrid models** can combine strengths of multiple models

Contributions

- We construct **WHiC**, a new dataset of hypernym pairs in context
  - Automatically extracted from **WordNet**!

Contributions

- We construct **WHiC**, a new dataset of hypernym pairs in context
  - Automatically extracted from **WordNet**!
- WHiC enables
  - **Empirical comparisons** of context embeddings
  - Analysis of **sensitivity to context,** and **directionality**

Contributions

- We construct **WHiC**, a new dataset of hypernym pairs in context
  - Automatically extracted from **WordNet**!
- WHiC enables
  - **Empirical comparisons** of context embeddings
  - Analysis of **sensitivity to context,** and **directionality**
- We provide experimental evidence of **challenging** nature of task

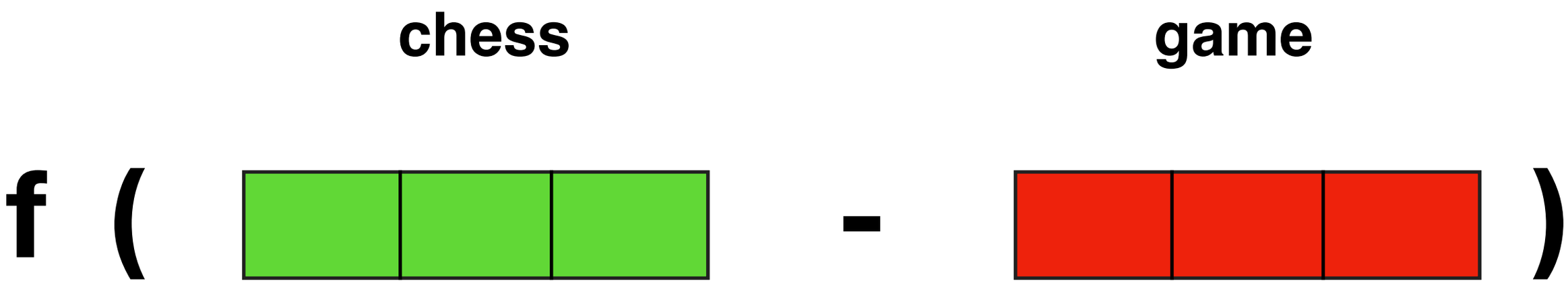
Contributions

- We construct **WHiC**, a new dataset of hypernym pairs in context
  - Automatically extracted from **WordNet**!
- WHiC enables
  - **Empirical comparisons** of context embeddings
  - Analysis of **sensitivity to context,** and **directionality**
- We provide experimental evidence of **challenging** nature of task

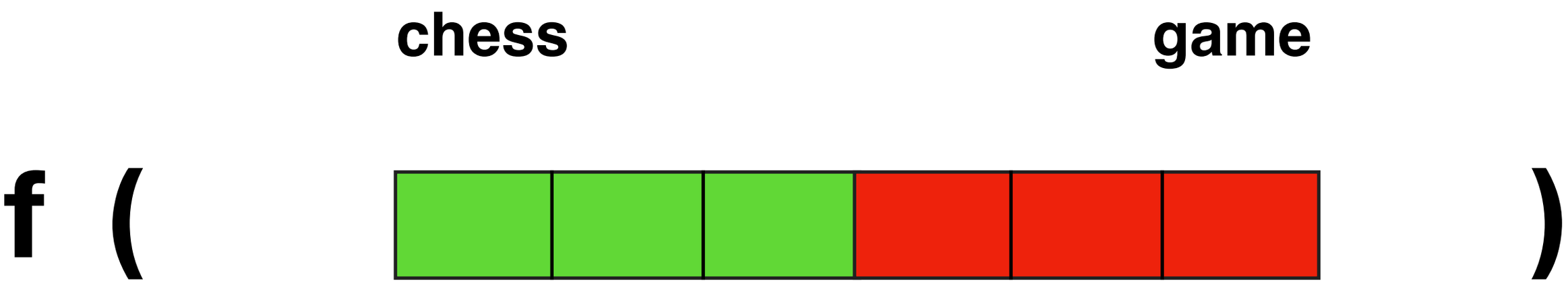
https://github.com/yogarshi/WHiC

Backup

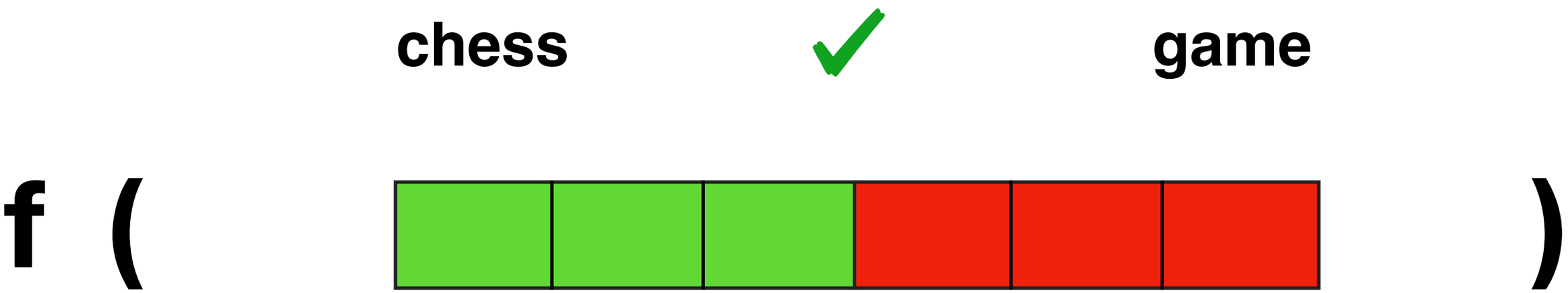
Supervised Hypernymy Detection



Supervised Hypernymy Detection

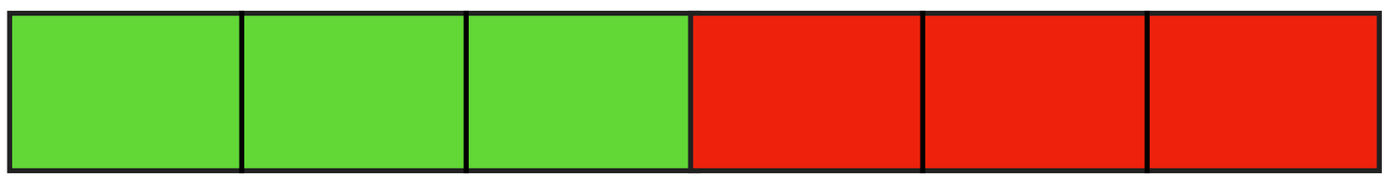


Supervised Hypernymy Detection



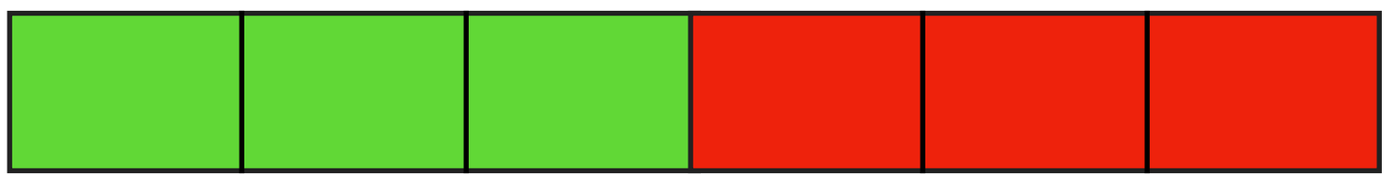
Supervised Hypernymy Detection

dog ✓ **game**

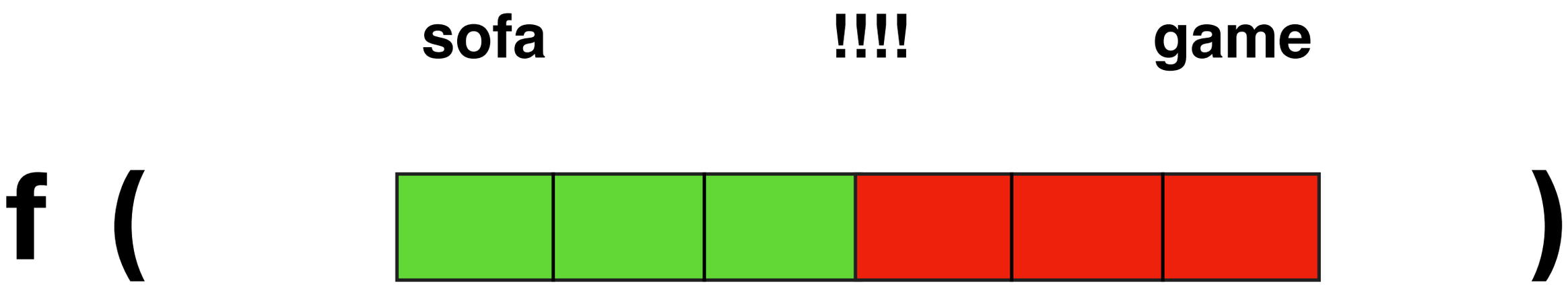
f ( **)**

Supervised Hypernymy Detection

sofa ✓ **game**

f ( **)**

Supervised Hypernymy Detection



H - Feature Detector

Roller and Erk; EMNLP 2016

H - Feature Detector

- Learn a separating hyperplane using a linear classifier

- Project word representations on the plane **orthogonal** to this hyperplane

H - Feature Detector

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

} Repeat

H - Feature Detector

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

} Repeat

Iteration *i* — words $\mathbf{x_i}$ and $\mathbf{y_i}$ — learned hyperplane $\mathbf{h_i}$

H - Feature Detector

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

} Repeat

Iteration *i* — words $\mathbf{x_i}$ and $\mathbf{y_i}$ — learned hyperplane $\mathbf{h_i}$

1. $\mathbf{x_i} \cdot \mathbf{y_i}$

H - Feature Detector

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

} Repeat

Iteration *i* — words $\mathbf{x_i}$ and $\mathbf{y_i}$ — learned hyperplane $\mathbf{h_i}$

1. $\mathbf{x_i} \cdot \mathbf{y_i}$
2. $\mathbf{x_i} \cdot \mathbf{h_i}$

H - Feature Detector

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

} Repeat

Iteration *i* — words $\mathbf{x_i}$ and $\mathbf{y_i}$ — learned hyperplane $\mathbf{h_i}$

1. $\mathbf{x_i} \cdot \mathbf{y_i}$
2. $\mathbf{x_i} \cdot \mathbf{h_i}$
3. $\mathbf{y_i} \cdot \mathbf{h_i}$

H - Feature Detector

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

} Repeat

Iteration *i* — words $\mathbf{x_i}$ and $\mathbf{y_i}$ — learned hyperplane $\mathbf{h_i}$

1. $\mathbf{x_i} \cdot \mathbf{y_i}$
2. $\mathbf{x_i} \cdot \mathbf{h_i}$
3. $\mathbf{y_i} \cdot \mathbf{h_i}$
4. $(\mathbf{y_i} - \mathbf{x_i}) \cdot \mathbf{h_i}$

H - Feature Detector

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

} Repeat

Iteration *i* — words $\mathbf{x_i}$ and $\mathbf{y_i}$ — learned hyperplane $\mathbf{h_i}$

1. $\mathbf{x_i} \cdot \mathbf{y_i}$
2. $\mathbf{x_i} \cdot \mathbf{h_i}$
3. $\mathbf{y_i} \cdot \mathbf{h_i}$
4. $(\mathbf{y_i} - \mathbf{x_i}) \cdot \mathbf{h_i}$

→ Non - linear classifier