Automatic recognition of habituals: a three-way classification of clausal aspect

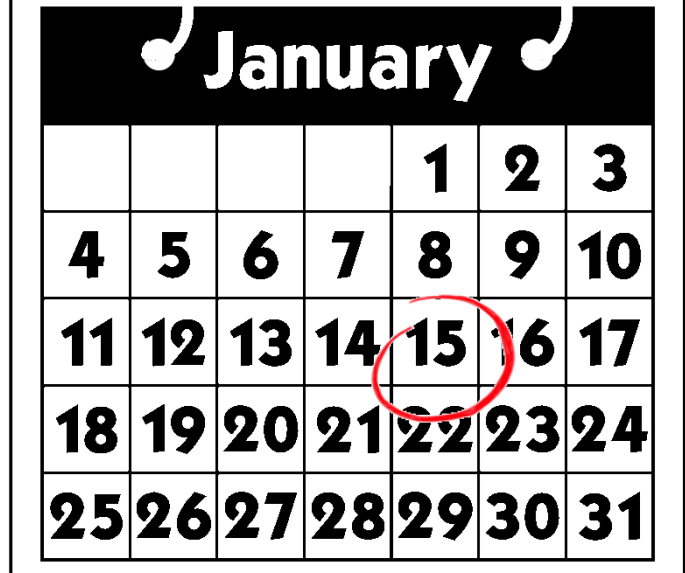**Annemarie Friedrich**   **Manfred Pinkal**

Department of Computational Linguistics, Saarland University

www.coli.uni-saarland.de/projects/sitent

Habituality

episodic: a particular event
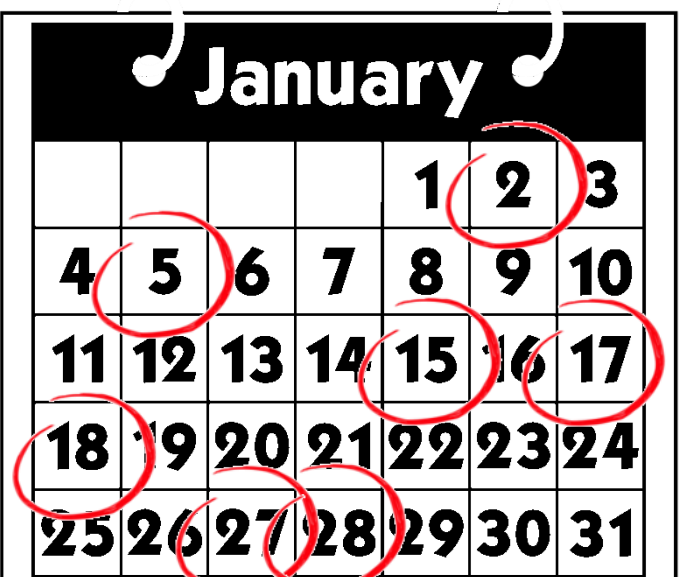**habitual**: generalization over situations, exceptions are tolerated

episodic
John went swimming yesterday!

Mathew & Katz (2009)

habitual
Bill often goes swimming.

Clausal aspect

lexical aspectual class + aspectual transformations (temporal) function of clause in discourse

similar: Xue & Zhang (2014)

clausal aspect		lexical aspect
episodic	Bill **drank** a coffee after lunch.	dynamic
habitual	Bill *usually* **drinks** coffee after lunch.	dynamic
	Italians **drink** coffee after lunch.	dynamic
	Sloths *sometimes* **sit** on top of branches.	stative
	John *never* **drinks** coffee.	dynamic
static	Bill **likes** coffee.	stative
	Bill *can* **swim**.	dynamic
	Bill *didn't* **drink** coffee yesterday.	dynamic
	Mary *has* **made** a cake.	dynamic

Lexical aspectual class

property of verb in context
dynamic: event, activity
  drink, swim, forget
stative: states, properties
  like, be, own

Siegel & McKeown (2000)
Zarcone & Lenci (2008)
Friedrich & Palmer (2014)

lexically stative clauses & clauses stativized via aspectual transformations such as negation, modals or English perfect
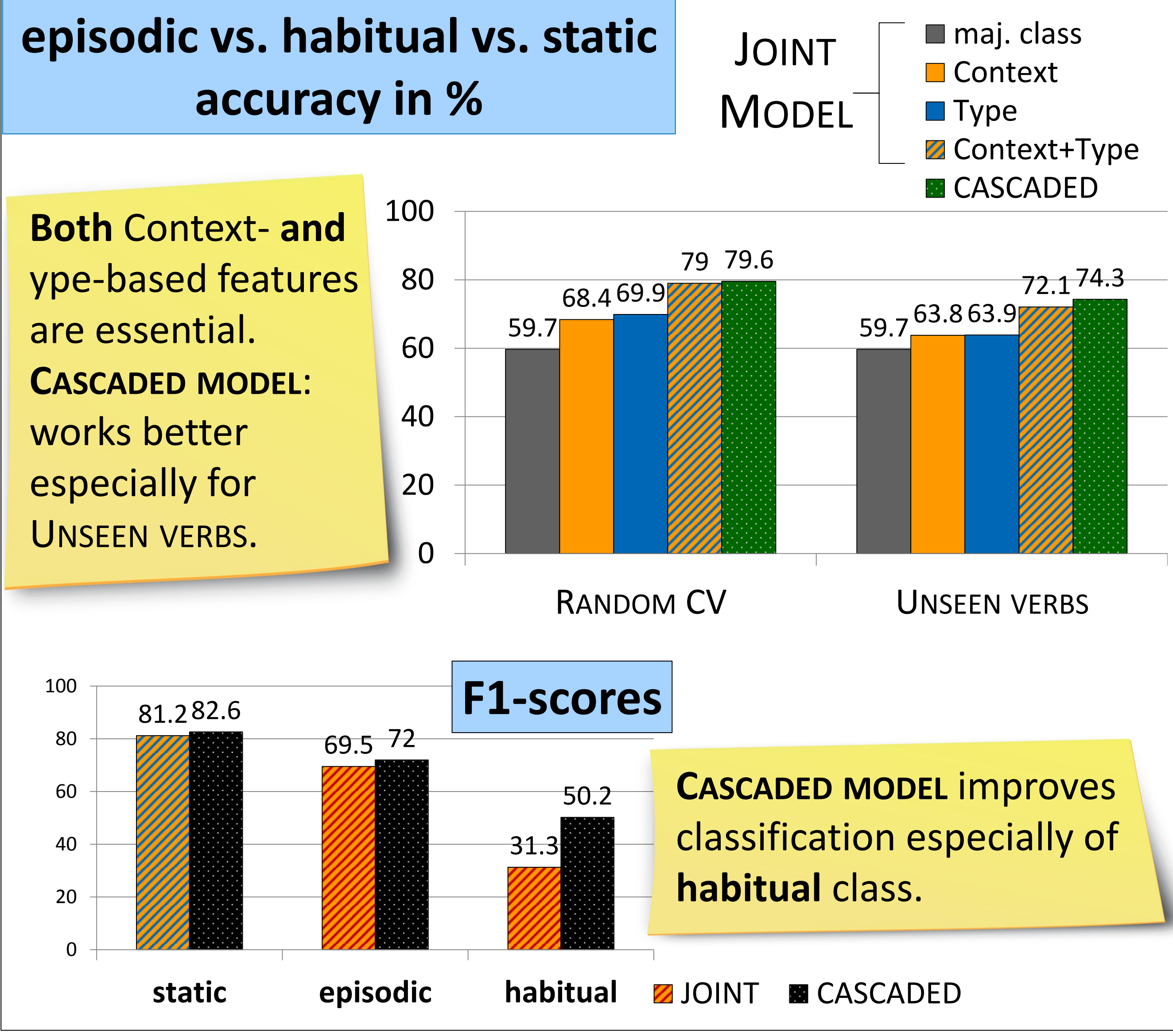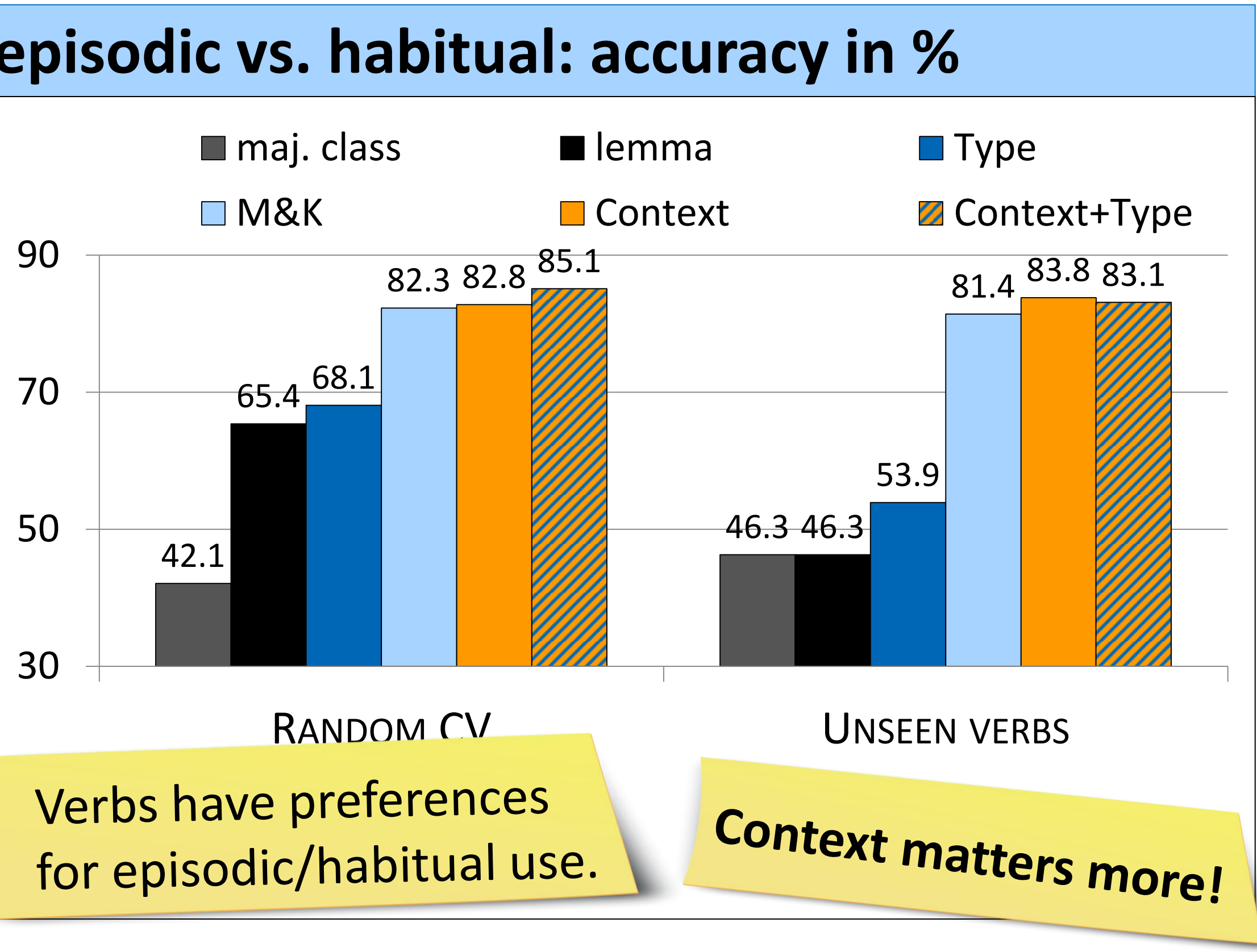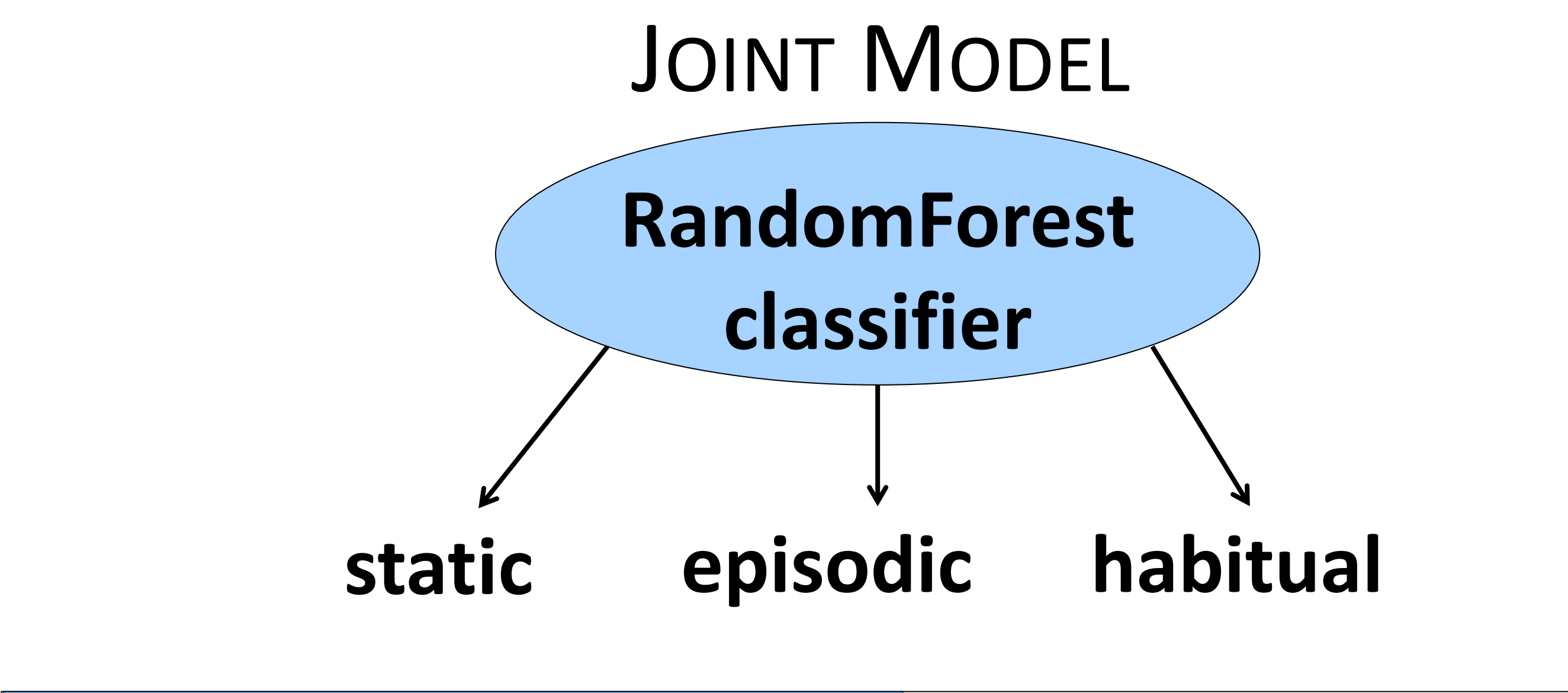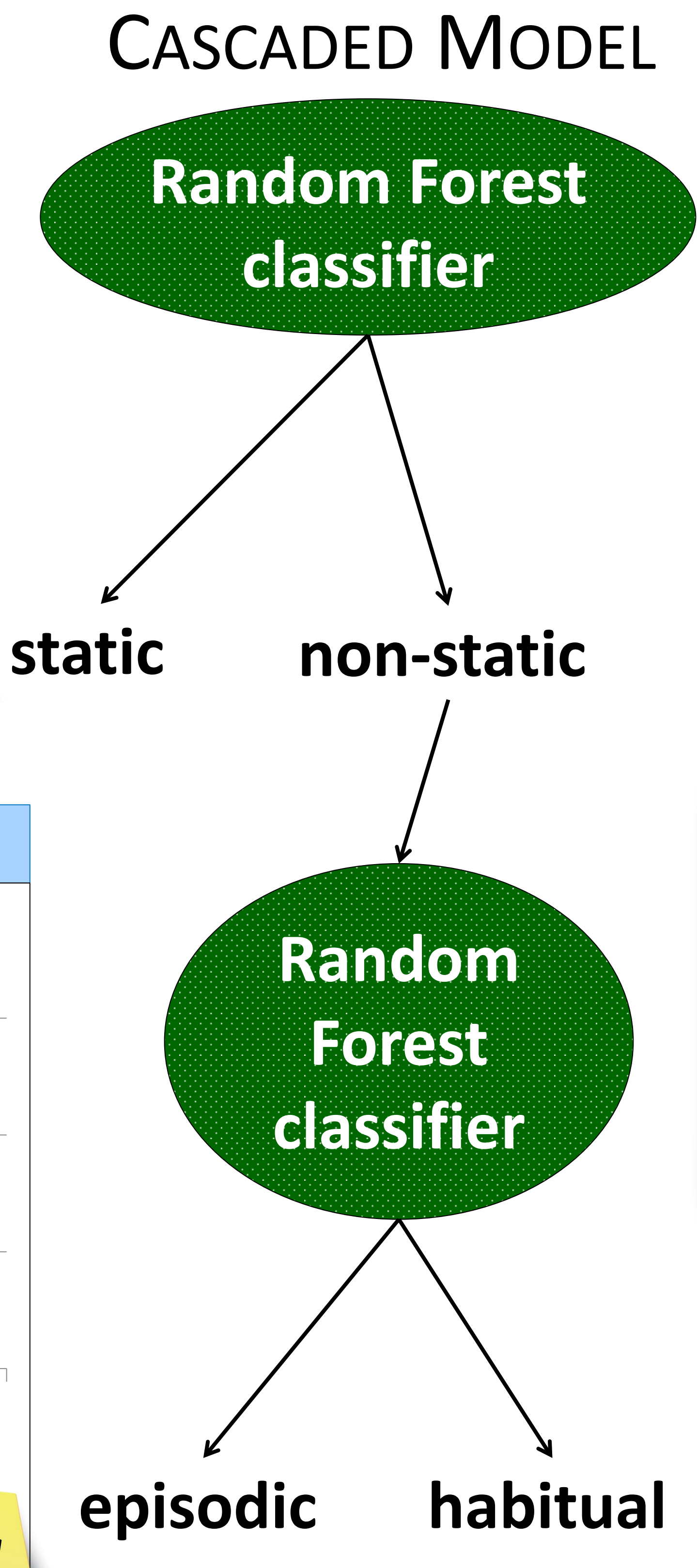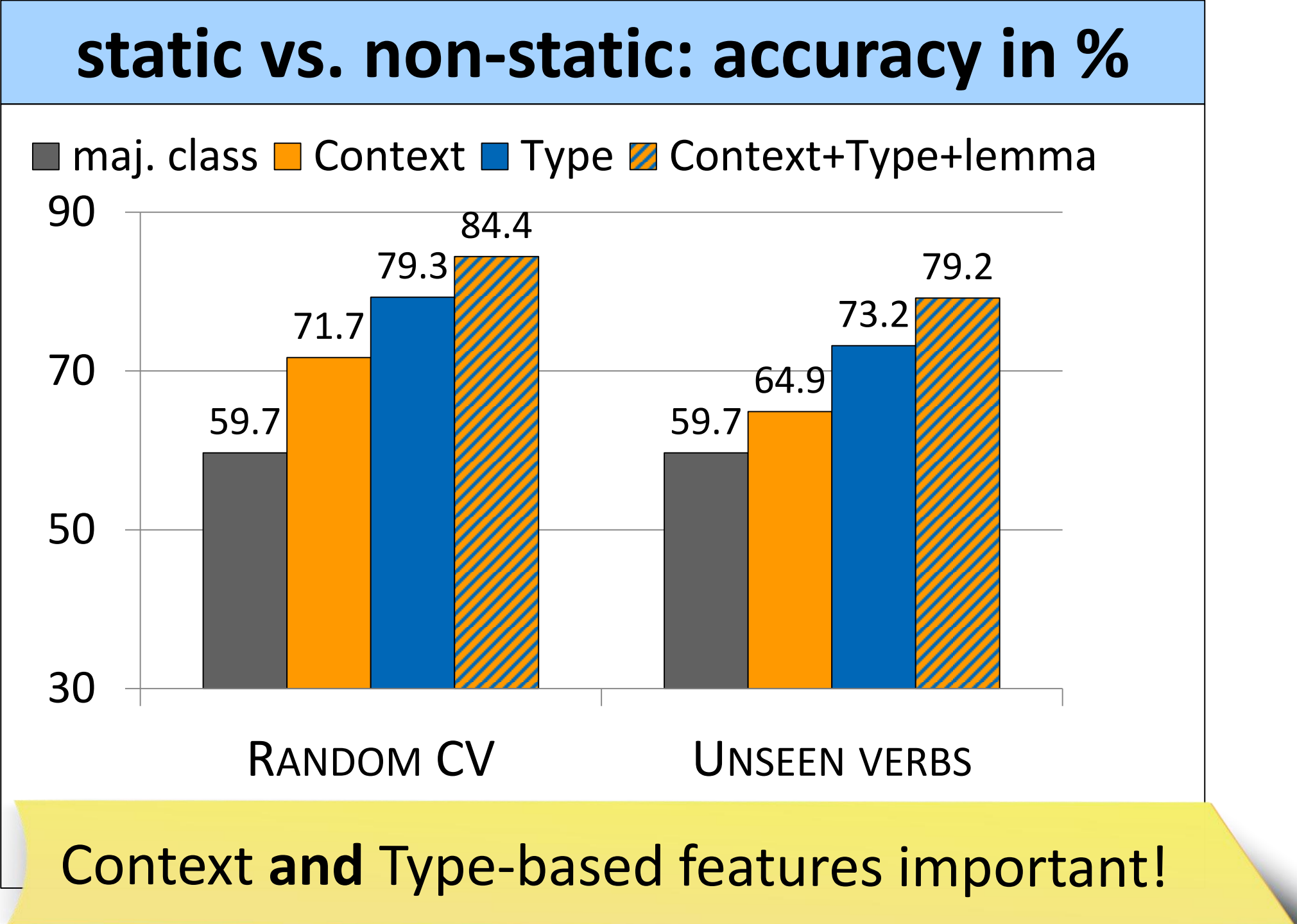
Future work: distinguish them!

Context-based features

verb: tense, POS, voice, progress., perfect
subject: bare plural, (in)definite
object: absent, bare plural, (in)definite
clause: modal, negated, conditional, tmod, …

John has **spilled** his coffee.
tense=past   perfect=true   modal=false

Type-based features – linguistic indicators

Siegel & McKeown (2000)

verb type: drink -- ling_ind_past = 0.0927
9.27% of all instances of *drink* in corpus are in past tense

frequency	negated	no subject
present	perfect	evaluation adverb
past	progressive	continuous adverb
future	for-PP	manner adverb
particle	in-PP	temporal adverb

parsed background corpus

Data

102 texts
10355 clauses

60% **static**
20% **episodic**
20% **habitual**

3 annotators, κ=0.61

static vs. non-static: accuracy in %

Legend: ■ maj. class ■ Context ■ Type ◪ Context+Type+lemma

RANDOM CV: 59.7, 71.7, 79.3, 84.4
UNSEEN VERBS: 59.7, 64.9, 73.2, 79.2

Context **and** Type-based features important!

episodic vs. habitual: accuracy in %

Legend: ■ maj. class ■ lemma ■ Type ■ M&K ■ Context ◪ Context+Type

RANDOM CV: 42.1, 65.4, 68.1, 82.3, 82.8, 85.1
UNSEEN VERBS: 46.3, 46.3, 53.9, 81.4, 83.8, 83.1

Verbs have preferences for episodic/habitual use.

Context matters more!

CASCADED MODEL

Random Forest classifier

static   non-static

Random Forest classifier

episodic   habitual

JOINT MODEL

RandomForest classifier

static   episodic   habitual

episodic vs. habitual vs. static accuracy in %

JOINT MODEL

Legend: ■ maj. class ■ Context ■ Type ◪ Context+Type ■ CASCADED

RANDOM CV: 59.7, 68.4, 69.9, 79, 79.6
UNSEEN VERBS: 59.7, 63.8, 63.9, 72.1, 74.3

Both Context- **and** Type-based features are essential.
CASCADED MODEL: works better especially for UNSEEN VERBS.

F1-scores

Legend: ◪ JOINT ■ CASCADED

static: 81.2, 82.6
episodic: 69.5, 72
habitual: 31.3, 50.2

CASCADED MODEL improves classification especially of **habitual** class.

Recognizing habituals and episodic clauses in free/complete text requires identification of lexically stative and other static clauses: **a three-way distinction is necessary**.

Best results achieved using sources of information relevant to lexical aspectual class **and** to other aspectual transformations.

Next steps

Leverage discourse context?
John rarely ate fruit. He just ate oranges.
Use aspectual distinctions to improve models of temporal discourse structure [Costa & Branco 2012]