Content Differences in Syntactic and Semantic Representations

Daniel Hershcovich & Omri Abend & Ari Rappoport

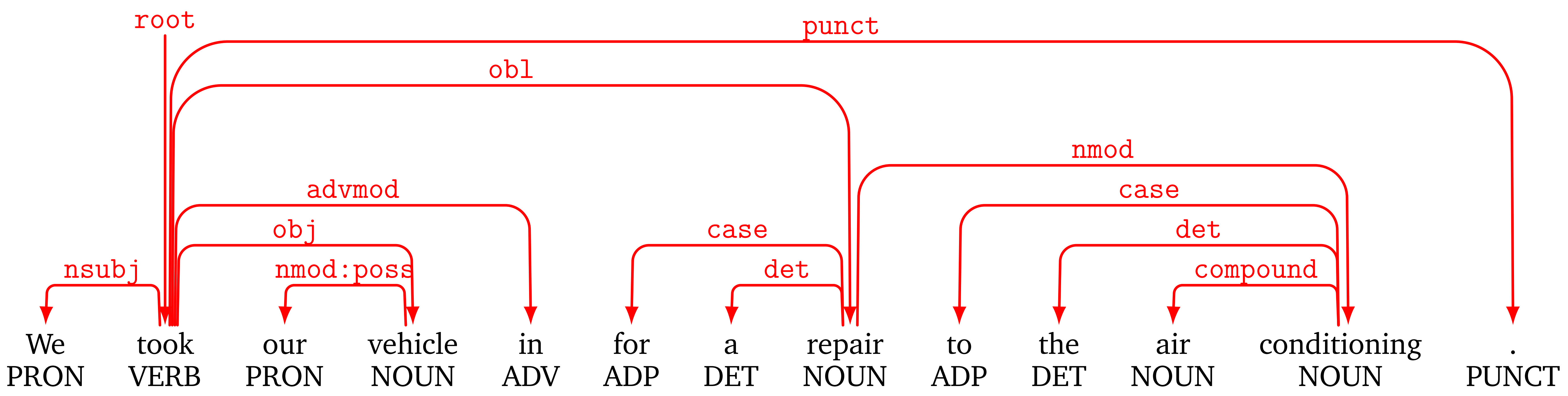
The Edmond & Lily Safra Center for Brain Sciences

ביה״ס להנדסה ולמדעי המחשב ע״ש רחל וסלים בנין
The Rachel and Selim Benin School of Computer Science and Engineering

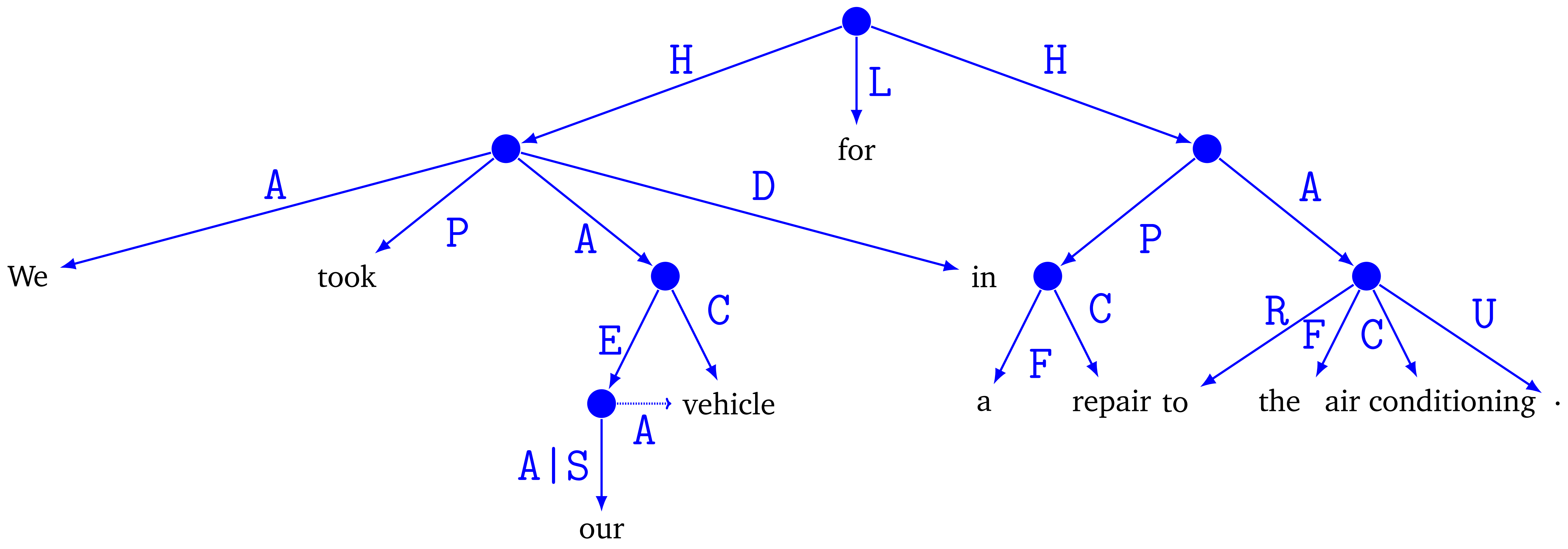
האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

How semantic are **Universal Dependencies (UD)**? We present a shared corpus and comparison with **Universal Conceptual Cognitive Annotation (UCCA)**

UD: cross-linguistically consistent grammatical dependency annotation [3].



UCCA: cross-lingual semantic representation. *Primary edges* form a tree. *Remotes* (dashed) allow reentrancy [1].



A	Participant
C	Center
D	Adverbial
E	Elaborator
F	Function
G	Ground
H	Parallel scene
L	Linker
P	Process
Q	Quantifier
R	Relator
S	State
T	Time
U	Punctuation

Shared Corpus

`https://github.com/UniversalConceptualCognitiveAnnotation/UCCA_English-EWT`

English Web Treebank (EWT) reviews section.

	A	A|P	A|S	C	D	E	F	G	H	L	N	P	Q	R	S	T	No Match
acl	58			1	4	249	1		48			6			1	1	409
advcl	14			12	2	2		6	512	4		11					423
advmod	225		1	69	1778	332	27	135	14	258	2	2	15	44	9	368	273
amod	25			134	647	837		1	28			7	130	3	269	25	176
appos	21			39	2	34			18						8		33
aux					384	2	1335			2		1		1			17
case	11			31	27	25	123			213	26	11	1	2629	154	1	262
cc				8	4	1	4	1	1	1567	381		6	12			52
ccomp	345			1		1			36			2			1	1	166
compound	225			116	67	586	21		2			32	19	1	12	24	683
conj	10			449	4	5		1	1262	1		6	2		10		497
cop				1			1312			1		9		10	178		7
csubj	13								3								46
det	10			17	119	440	2963				1		129	16	1		124
discourse	1			2	1		25	29	27	16					5		19
expl	21			1			98								17		3
iobj	131			1			1										10
list	3			7	2	1			27						1		6
mark				9	7	1	531	1		654				407	1	5	143
nmod	844	1	1	20	9	786	8	4	12	1	1	20	2	2	11	27	488
nsubj	4296	7	21	25	3	2	55	1	5	61		58	1	80	14	4	247
nummod	2			33	12	17		4		4			334				64
obj	1845		1	54	21	6	11	1	4	23		52	1	23	3	11	583
obl	1195			19	115	41	1	17	39	34		6	6	26	7	302	611
parataxis	6		1	5		4		6	285						3		180
vocative	17							8									
xcomp	121			4	25				8			38			38		526
head	445	48	159	6388	717	142	564	83	2462	42	1	4163	120	52	1547	32	2235
No Match	1421	37	58	640	417	291	14	33	2291	146	6	802	94	52	369	96	

	Train	Dev	Test
# Passages	347	192	184
# Sentences	2,723	554	535
# Tokens	44,804	5,394	5,381

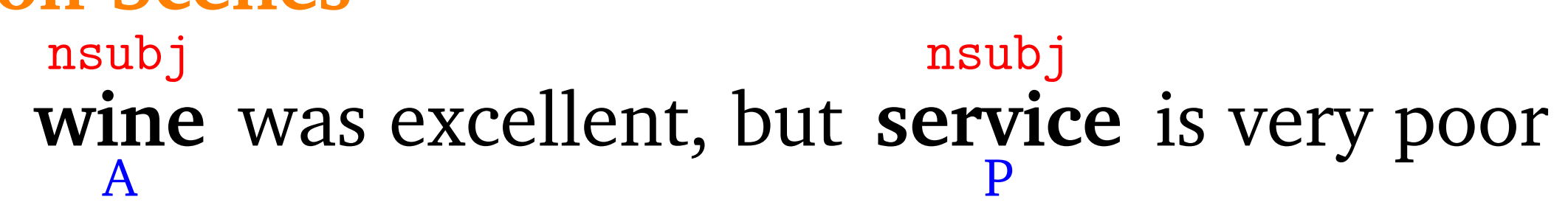
UCCA parsing results with TUPA [2]:

Features	Primary LP	Primary LR	Primary LF	Remote LP	Remote LR	Remote LF
Dev						
Gold UD	72.1	71.2	71.7	61.2	38.1	47.0
UDPipe	73.0	72.1	72.5	53.7	40.8	46.4
StanfordNLP	73.7	72.7	73.2	52.3	39.2	44.8
Test						
Gold UD	72.2	71.2	71.7	60.9	36.8	45.9
UDPipe	72.4	71.7	72.1	60.3	38.5	47.0
StanfordNLP	72.7	72.4	72.6	59.6	35.4	44.4

Comparison

bit.ly/uccaud

Scenes/Non-Scenes

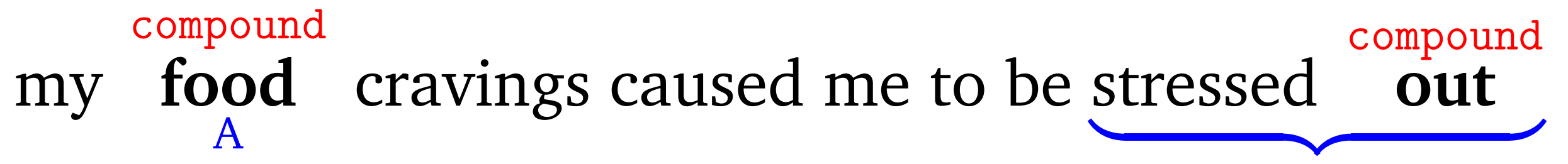


wine was excellent, but **service** is very poor

Primary/Secondary Relations

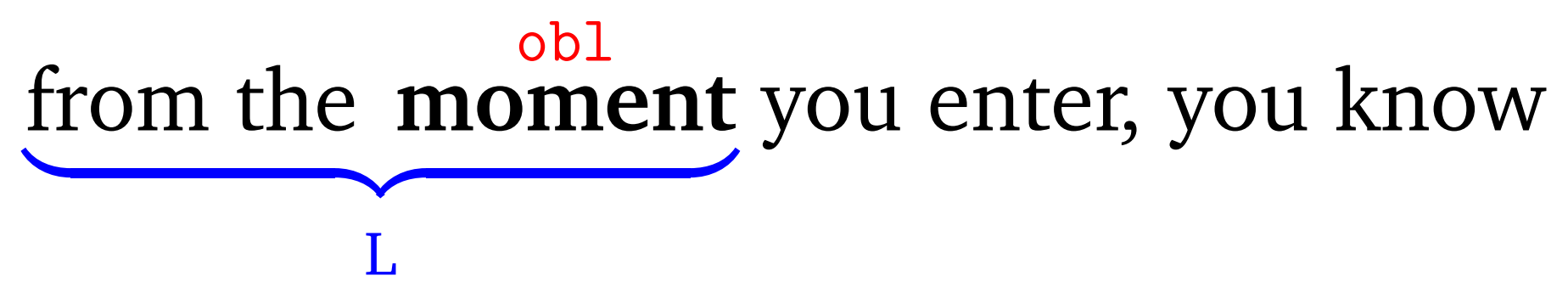
I will never come **here again**

Multi-Word Expressions



my **food** cravings caused me to be stressed **out**

Linkage



from the **moment** you enter, you know

[1] Omri Abend and Ari Rappoport. Universal Conceptual Cognitive Annotation (UCCA). In *Proc. of ACL*, pages 228–238, August 2013.

[2] Daniel Hershcovich, Omri Abend, and Ari Rappoport. A transition-based directed acyclic graph parser for UCCA. In *Proc. of ACL*, pages 1127–1138, 2017.

[3] Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. Universal dependencies v1: A multilingual treebank collection. In *Proc. of LREC*, May 2016.

Please participate in the CoNLL 2019 Shared Task:
Cross-Framework Meaning Representation Parsing
SDP, EDS, AMR and UCCA mrp.nlpl.eu