

We pilots

must-see television

London, UK

Lake Michigan

King Henry VIII

French actor Mr. Gaspard Ulliel

July 31, 1980

one thousand, one
hundred and one point one

I myself

Figure 4

Symphony No. 5

\$15 an hour

Mischievous Nominal Constructions in Universal Dependencies

Nathan Schneider, Amir Zeldes (Georgetown University)

UDW @ SyntaxFest 2021 • March 23, 2022 • Online

Mischievous Nominal Constructions

- Various kinds of nominal phrases don't "fit the mold" of current UD relations
 - Certain compounds, name expressions, date/time/quantity expressions
- This paper identifies challenges and proposes solutions/alternatives
 - Goal: spark discussion → better guidelines → harmonization across treebanks
 - Paper focuses on English (though ideally many of the improvements will be universal)
- Let's do this interactively!

Problem 1: Complex Personal Names

Problem 2: Other Complex Names

Problem 3: Dates

Problem 4: Phrasal Compounds

Problem 1: Complex Personal Names

He may be more commonly known (in the UK at least) as “that bloke from the Chanel advert”, but French actor Mr Gaspard Ulliel is clearly no one-trick pony. (mrporter.com)

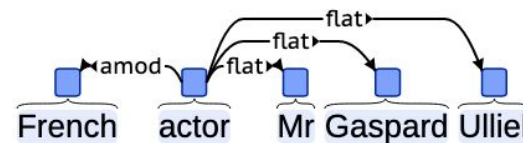
VOTE »

In “**French actor Mr Gaspard Ulliel**”, which word should be the head?

Problem 1: Complex Personal Names

He may be more commonly known (in the UK at least) as “that bloke from the Chanel advert”, but French actor Mr Gaspard Ulliel is clearly no one-trick pony. (mrporter.com)

- Currently subsumed under flat
 - “In many languages, there are multiword proper names with no clear internal syntactic structure and no clear evidence that one of the words is the syntactic head. Such names are annotated using the flat relation...Titles/honorifics are also analyzed using the flat relation.” ([guidelines](#))

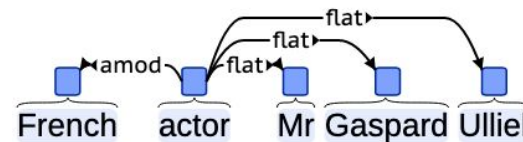


Problem 1: Complex Personal Names

He may be more commonly known (in the UK at least) as “that bloke from the Chanel advert”, but French actor Mr Gaspard Ulliel is clearly no one-trick pony. (mrporter.com)

- Currently subsumed under flat

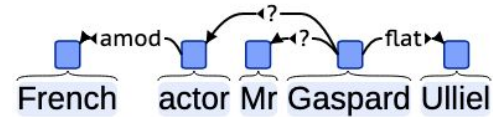
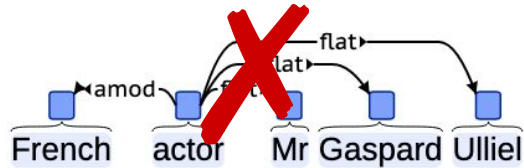
- “In many languages, there are multiword proper names with no clear internal syntactic structure and no clear evidence that one of the words is the syntactic head. Such names are annotated using the flat relation...Titles/honorifics are also analyzed using the flat relation.” ([guidelines](#))



- Evidence that “French actor” and “Mr” are **modifiers** of the main part of the name, “Gaspard Ulliel”:
 - “actor” lacks a determiner: *French actor is clearly no one-trick pony.
 - Optionality: (French actor) (Mr) Gaspard Ulliel is clearly no one-trick pony.
- Also: official titles, e.g. “Secretary of State Hillary Clinton”

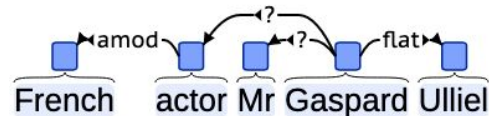
Problem 1: Complex Personal Names

- Proposed structure: titles/honorifics/descriptors are modifiers!



Problem 1: Complex Personal Names

- What deprel to use for these modifiers?

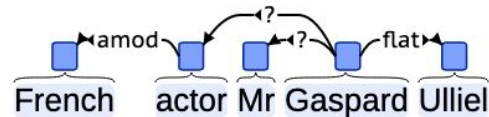


- ➡ A **new relation** might be cleanest, but is a long-term solution—cannot be added until UDv3.
- ➡ **appos** corresponds to identity semantics. But usually applies to full NPs (“French actor” would be expected to have a determiner) and order is reversible (*Gaspard Ulliel, Mr, ...)
- ➡ Nouns typically premodify other nouns as **compound**. But usually singular (*eggs cartons); here there is number agreement: French actors Ulliel and Marceau
- ➡ **nmod** is most generic for nominals-modifying-nominals. But English **nmod** usually marked with a preposition (as **case**).

VOTE » Closest deprel for “actor” and “Mr”: **appos, compound, nmod, other?**

Problem 1: Complex Personal Names

- What deprel to use for these modifiers?



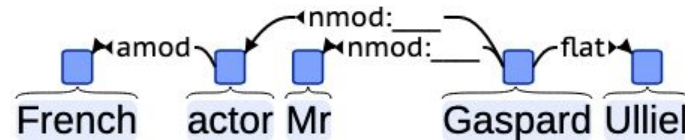
Is it a kind of compound? appos? nmod? A new relation entirely?

- Pros/cons...

- ☞ A **new relation** is impractical—cannot be added until UDv3.
- ☞ **appos** usually applies to full NPs (“French actor” would be expected to have a determiner) and order is reversible (*Gaspard Ulliel, Mr, ...)
- ☞ English **compound** modifiers usually singular (*eggs cartons), but here there is number agreement: French actors Ulliel and Marceau
- ☞ Plain **nmod** in English is normally marked with a preposition (as case)
- ☞ **Proposed solution: a new nmod subtype**

Problem 1: Complex Personal Names

- Proposed solution: a new **nmod** subtype
- What to name the subtype?



- ☞ **nmod:title** because these constructions involve titles in personal names, or
- ☞ **nmod:desc** for “descriptor”. This is a broader term that could include other mischievous nominal modifiers, e.g. nominal person (cf. Höhn, earlier today):
 - We pilots/You guys deserve a raise
 - not appos: *Pilots we deserve a raise
 - *We* as head of *pilots*, because omitting it changes the meaning significantly
 - cf. pilots Earhart and Lindbergh

Problem 2: Other Complex Names

- UD policy: **syntactically regular named entities** do not get special treatment (except NOUN vs. PROPN)
- “Church Street”: standard right-headed N-N compound
- “Lake Michigan”, “Figure 4” are tricky cases

VOTE »

Which word is the head in “**Lake Michigan**”? In “**Figure 4**”?

Problem 2: Other Complex Names

- “Lake Michigan”: which word is the head?

- Lake Michigan is a lake, not a Michigan

- **Plural:** Lakes Michigan and Ontario cf.: Church and Broad Streets

- **Both words essential to the proper name:**

I went to Lake Michigan → I went to the lake; *I went to Lake; *I went to Michigan

- **Head-first order perhaps due to borrowing of place names from Romance languages**



Problem 2: Other Complex Names

- “Lake Michigan”: which word is the head?

- Lake Michigan is a lake, not a Michigan
- Plural: Lakes Michigan and Ontario cf.: Church and Broad Streets
- Both words essential to the proper name:
I went to Lake Michigan → I went to the lake; *I went to Lake; *I went to Michigan
- Head-first order perhaps due to borrowing of place names from Romance languages



- Numbered entities

- “Figure 4”: morphosyntactically similar to “Lake Michigan” but with a number as designator. Is the number a compound modifier? **nmod:desc**? Add **nummod:name**? (We couldn’t decide!)
- “Item number 4”: “number” is optional & not typically pluralized in this construction—avoiding the plural is characteristic of compound modifiers:
items 4 and 5; items number 4 and 5

☞ “number” as compound modifier of “4”

	head	modifier optional?	invertible?	agreement?	type	relation
(2a) actor Ulliel	R	Ulliel	*Ulliel, actor	actors Ulliel and Marceau	name (head)	← nmod:desc
§2 President Obama	R	Obama	*Obama, President	Presidents Obama and Biden	name	← nmod:desc
§3.1 Church Street	R	*Street / the street	*Street, Church	Church and River Streets	name	← compound
(12) Lake Michigan	L	*Lake / the lake	*Michigan, Lake	Lakes Michigan and Ontario	name	compound →
(14) Figure 4	L	*Figure / the figure	*4, Figure	Figures 4 and 5	name w/ num	nummod:name →? compound →? nmod:desc →?
(13) Firefox 58.0	L	Firefox	*58.0, Firefox	*Firefoxes 58.0 and 59.0	name w/ num	nummod:name →? flat ?
§3.7 London, UK	L	London	*UK, London	*Londons, UK and Ontario	name	nmod:npm →
Joe Biden	-	(flat)	(flat)	*Joe and Jill Bidens	name	flat
(6) my brother Sam	L	my brother	Sam, my brother	my brothers Sam and John	name (mod)	appos →

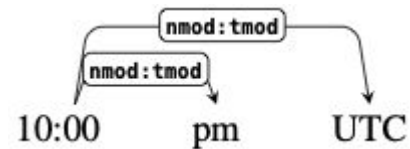
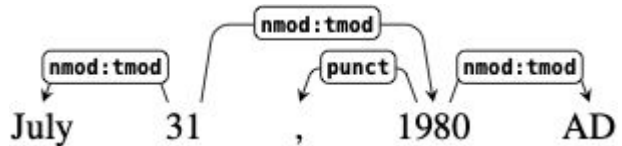
Table 1: Constructions involving names and their syntactic properties.

Problem 3: Dates

- Currently no guidance for dates and times, where headedness is not obvious.
We propose:

☞ **Day as head** of month and year (cf. proposal by Dan Zeman, this workshop!)

☞ **nmod:tmod** subtype to relate the parts of a date expression



Problem 4: Phrasal Compounds

	Inner relation(s)	
a <u>high - quality</u> product		
a <u>by - the - book</u> strategy		
a <u>must - see</u> movie		
a <u>fly - by - night</u> operation		
a <u>10 - year</u> plan		
a <u>cost - effective</u> approach		
<u>fire - breathing</u> dragons		

Problem 4: Phrasal Compounds

	Inner relation(s)	
	☞ transparent syntactic relations if possible; compound by default	
a <u>high - quality</u> product	amod	
a <u>by - the - book</u> strategy	case, det	
a <u>must - see</u> movie	aux	
a <u>fly - by - night</u> operation	case, obl	
a <u>10 - year</u> plan		
a <u>cost - effective</u> approach		
<u>fire - breathing</u> dragons		

Problem 4: Phrasal Compounds

	Inner relation(s)	
	☞ transparent syntactic relations if possible; compound by default	
a <u>high - quality</u> product	amod	
a <u>by - the - book</u> strategy	case, det	
a <u>must - see</u> movie	aux	
a <u>fly - by - night</u> operation	case, obl	
a <u>10 - year</u> plan	compound (not nummod)	
a <u>cost - effective</u> approach	compound	
<u>fire - breathing</u> dragons	compound (not obj)	

Problem 4: Phrasal Compounds

	Inner relation(s)	Outer relation
	☞ transparent syntactic relations if possible; compound by default	☞ amod if ADJ/participle, else compound
a <u>high - quality</u> product	amod	compound
a <u>by - the - book</u> strategy	case, det	compound (not nmod)
a <u>must - see</u> movie	aux	compound
a <u>fly - by - night</u> operation	case, obj	compound
a <u>10 - year</u> plan	compound (not nummod)	compound (not nmod : tmod)
a <u>cost - effective</u> approach	compound	amod
<u>fire - breathing</u> dragons	compound (not obj)	amod

Thanks!

- Extended version of the paper: <https://arxiv.org/abs/2108.12928>

Want to contribute to English UD?

- Volunteers welcome!
- A good way to get involved is by following the the issue tracker (https://github.com/UniversalDependencies/UD_English-EWT/issues) and participating in discussions.