



HIPE 2022 Evaluation Lab

Named Entity Processing in Multilingual Historical Documents

Maud Ehrmann (EPFL), Matteo Romanello (UNIL),
Sven Najem-Meyer (EPFL), Antoine Doucet (ULR), Simon Clematide (UZH)

CLEF 2022 (hybrid), 07.09.2022



HIPE 2022 in a nutshell

- Identifying Historical People, Places and other Entities.
- Second shared task on Named Entity Processing in Historical Documents.
- NE Recognition and Classification & NE Linking.
- Participating teams: 5

The screenshot shows the HIPE 2022 website. At the top, there is a navigation bar with links for "HIPE 2022", "About", "Tasks & Data", "Evaluation", "Results", "Timeline", "Workshop", "References", and a menu icon. Below the navigation is a large banner with the text "HIPE – Identifying Historical People, Places and other Entities" and a subtitle "Shared Task on Named Entity Recognition and Linking in Multilingual Historical Documents". The main content area is divided into several sections: "About HIPE-2022" (with a "Learn more" link), "Tasks & Data" (with a "Learn more" link), "Timeline & Instructions" (with a "Learn more" link), "Evaluation" (with a "Learn more" link), "Workshop" (with a "Learn more" link), and "Organizers" (with a "Learn more" link). The background of the main content area features a grayscale image of historical documents.

<https://hipe-eval.github.io/HIPE-2022/>



Why HIPE?

New data

Emergence of large-scale
archives of digitized
contents



New needs

Content retrieval by
humanities scholars



Application example: Semantic indexing of historical newspapers

The screenshot shows the impresso Media Monitoring of the Past web application. The main search bar at the top contains the query "otto von Bism". Below the search bar, a dropdown menu lists several search results:

- ...otto von Bism (Search in article contents)
- ...otto von Bism (Search in article titles)
- Otto von Bismarck
- prince Otto von Bismarck
- Otto von Bismark (highlighted in blue)
- Otto von Bismarck
- Otto Christian Archibald von Bismarck
- ... otto von Bism (More Persons ...)

Below the search results, there is information about the number of publications and named entities:

- 2 countries of publication
- 530,086 named entities disambiguated

Links for contacting the project and social media are provided:

- More? Check on our [blog](#)
- info @ impresso-project [dot] ch
- project website: impresso-project.ch
- github: [impresso](#)
- twitter: [@impressoProject](#)

At the bottom left, there are buttons for "LINES: OFF" and "DARK MODE: ON".

The main content area features a large title "Media Monitoring of the Past" in gold-colored serif font. Below the title, a subtext reads "Mining 200 years of historical newspapers" with a small info icon. A question "How can newspapers help understand the past? How to explore them?" is also present.

A yellow banner at the bottom left contains the text "Searching named entities (among others) over millions of articles".

At the bottom right, there is a note about legal access: "For legal reasons not all content is available in Open Access. To gain full access: [LOAD NON-DISCLOSURE-AGREEMENT FORM](#) ... and return the signed form to info@impresso-project.ch".

Searching named entities (among
others) over millions of articles

TABLE OF CONTENTS

43 articles in 4 pages (Personal use (no export))

Add filters from your current search query

* 3 search filters can't be applied.

Search words...

show only matching articles (no results)

La Journée politique

M. de Brazza est arrivé lundi à bord du Stamboul, courrier de la côte occidentale d'Afrique. M. de Brazza est parti hier pour Alg [article] short text p.1

Continue reading: TRANSCRIPT

LOCATIONS France, Congo Free State, Westminster, Rome, Naples, Italy, Bulgari, Russia
PEOPLE Pierre Savorgnan de Brazza, John Redmond, Ferdinand I of Bulgaria

A la Chambre Séance du 12 février
Présidence de M. de Wacquani,
présid[...]

A la Chambre Séance du 12 février Présidence de M. de Wacquani, président. Décidément certains de nos députés trouvent qu'ils ont le temps pour s'occ [article] short text pp.2,3 (2 pages)

Continue reading: TRANSCRIPT

LOCATIONS Laval, Mayenne
PEOPLE Victor de Tornaco

Sans titre

Postas. — Un député de notre ville a tonné il y a

L'indépendance luxembourgeoise · Wednesday, February 13, 1895 [4] p.3 of 4



Faux départ

LOCATIONS Marcel Dallermagne

PEOPLE Otto von Bismarck, Guillaume Cale

TOPICS

42.5% FR gouvernement - parti - ministre - président - politique

9.2% FR vie - esprit - pays - travail - effort

18.6% FR roi - prince - empereur - comte - reine

12.2% FR question - point - pays - conseil - gouvernement

ADD TO COLLECTION ...

NO TEXT REUSE
PASAGES AVAILABLE



L'opinion publique
assez vivement pré-
de politique qui im-
en dépit des déme

gement de person-
l'empire. Elle note
symptômes qui lui permettent de porter
un jugement sur les tendances actuelle-
ment prévalantes en haut lieu. Les dé-
bats du Reichstag, ceux surtout de la
commission chargée d'examiner le projet
contre les menées subversives, ont déjà
jeté un jour assez vif sur ce point.

un jugement sur les tendances
actuellement prévalentes en haut lieu. Les
débats du Reichstag, ceux surtout de la
commission chargée d'examiner le projet
contre les menées subversives, ont déjà
jeté un jour assez vif sur ce point.
Bien que le prince de Hohenlohe,
fidèle aux traditions d'une vie presque
tout entière passée dans la diplomatie,

Visualising facsimile, OCR
and entity mentions.



Otto von Bismarck PERSON

OVERVIEW

317 RELATED ARTICLES

MENTIONED 334 TIMES

OPEN IN SEARCH PAGE...

"German statesman and Chancellor (1815-1898)" (wikidata)



W

Otto von Bismarck

German statesman and Chancellor (1815-1898)

Schönhausen, 1815 - Friedrichsruh, 1898

SOURCE: W/Q8442



COUNTRY

Switzerland ⊕

LANGUAGE

German ⊕

TYPE

article ⊕

244

Luxembourg ⊕

French ⊕

page article ⊕

66

unclassified content ⊕

5

=

2

advertisement ⊕

PERSON

Otto von Bismarck ⊕

LOCATION

Berlin ⊕

TOPIC

DE regierung · paris · frankreich · minister · kammer ⊕

113

Wilhelm II, German Emperor ⊕

Germany ⊕

FR roi · prince · empereur · comte · reine ⊕

87

Herbert von Bismarck ⊕

Auch ⊕

DE könig · kaiser · königin · prinz · prinzessin ⊕

68

Karl Marx ⊕

Paris ⊕

DE verlag · buch · band · geschichte · werk ⊕

62

Adolf Hitler ⊕

Switzerland ⊕

DE krieg · deutschland · frankreich · welt · friede ⊕

59

Richard Wagner ⊕

France ⊕

DE welt · leben · mensch · wort · art ⊕

43

Debbie Reynolds ⊕

England ⊕

DE mann · hand · kopf · nacht · gesicht ⊕

42

François Duviller ⊕

Lage ⊕

FR gouvernement · parti · ministre · président · politique ⊕

42

Frederick III, German Emperor ⊕

Marcel Dallermagne ⊕

FR guerre · paix · pays · peuple · politique ⊕

40

Frederick II of Prussia ⊕

Italy ⊕

FR vie · monde · mort · fol · peuple ⊕

39

PARTNER

ACCESSRIGHT

COLLECTION

157

Personal use ⊕

0

1

>

2

Personal use (no export) ⊕

80

Overview of a NE, with e.g.
related persons and topics.



Why HIPE?

New data

Emergence of large-scale
archives of digitized
contents



New needs

Content retrieval by
humanities scholars



Challenges

NLP on historical texts is
hard

Spelling variations
Noisy OCR
Data sparsity
Limited resources
Multilingualism
Variety of documents

HIPE 2020

Objectives

- Strengthen the **robustness** of approaches;
- Enable **performance comparison**;
- Foster **efficient semantic indexing** of digitized cultural heritage collections.

Data Historical newspapers in Fr, De, En from 1738 to 2019.

Outcomes

- **Neural-based systems** with strong resources can achieve good NERC performances;
- Performances are affected by OCR noise, but less by document publication date;
- Progress needed to better handle **OCR noise, nested entities, small-data settings, and entity linking**.



HIPE 2022

Novelties

- Addition of a **new document type**: historical newspapers + classical commentaries;
- Consideration of a **broader language spectrum** (5 for newspapers, 3 for commentaries);
- Confrontation with **heterogeneous annotation tag sets and guidelines**.

Objectives

- Strengthen the **robustness** of approaches;
- Enable **performance comparison**;
- Advance the **transferability** of NE processing approaches **across languages, time periods, document and annotation types**;
- Foster **efficient semantic indexing** of cultural heritage collections.

Tasks

- **Named Entity Recognition and Classification (NERC)**
 - NERC-Coarse: recognition and classification of high-level entity types.
 - NERC-Fine: recognition and classification of fine-grained entity types + nested entities (En, Fr, De).
- **Named Entity Linking (EL)**
 - Linking of NE mentions to a unique item ID in Wikidata or to a NIL value if the mention does not have a corresponding item in the KB.

HIPE-2022 Data

Six NE-annotated datasets composed of **historical newspapers** and **classical commentaries** covering ca. 200 years.

Dataset	Doc. type	Language	Project	Suitable for
hipe2020	newspapers	Fr, De, En	impresso	NERC-Coarse, NERC-Fine, EL
newseye	newspapers	Fr, De, Sw, Fi	NewsEye	NERC-Coarse, NERC-Fine, EL
sonar	newspapers	De	SoNAR (SBB)	NERC-Coarse, EL
letemps	newspapers	Fr	-	NERC-Coarse
topres19th	newspapers	En	Living with Machines	NERC-Coarse, EL
ajmc	commentaries	Fr, De, En	AjMC	NERC-Coarse, NERC-Fine, EL



ΣΟΦΟΚΛΕΟΥΣ

ΧΟ. ἀνὴρ φρονεῖ ξουκεν. δᾶλ' ἀνοίγετε.
τάχι ἀν τοῦ αἰδῶ κατ' ἐμοὶ βλέπεται λάβοι.
ΤΕ. Ιδού, διοίγω· προσθήτεν δὲ ξεστοῖ σοι
τὰ τούθε πράγη, καντός ὡς ἔχων κυρεῖ.

στρ. 4. AI. ἴω
2 φίλοι ναυπάταια, μόνοι ἔχον φίλων,
3 μίνοι ἔτι ἐμμένοντες ὄρθῳ νόμῳ,
4 ιδεοθέ μ' οἷον ἀρτοί κίμη φωνίας ὑπὸ ζαλῆς
5 δρμφίσμονος κυκλεῖται.

ΧΟ. οἵμ' ἦς ξουκα ὄρθῳ μαρτυρεῖ ἄγαν.
δηλοῖ δὲ τούρην ως ἀφροτοσιστος ζεῖται.

ἀττ. 5. AI. ἴω
2 γένες ύλας ἀρπάγει τέχνας.

664 L has the *s* of *τέχνα* from a later hand.—*ἀρπάγει* Wecklein writes *ἀρπάγει* &
665 *ετεῖ* made in L from *χεῖται*: this is explained by the false reading *χεῖται* in Pal.—*τέχνας* conj. *τεῖται* *φέρει*.

Ajax, whose mother was Ethobea (σφέα).—*τεῖται*, *χεῖται*: the phrase *τεῖται* *χεῖται* is probably a corruption of *τεῖται* *χεῖται* *φέρει*. The use of *τεῖται* from *χεῖται* deserves notice, as suggesting the possibility that *τεῖται* *ετεῖ* may mean 'has been born' (which *χεῖται* *ετεῖ* does) even if it is a usage which, however, lacks proof: see on *EL* 1975.—*λαρυγνεῖς*: he had gone to the *EL* 1975.—*λαρυγνεῖς* *τραγεῖς* (*τραγεῖς* *ετεῖ*) to be despised. Cf. Thuc. i. 11 § 1 (the Greeks at Troy) *πατεῖσθαι...τραγεῖς* *τραγεῖς* *την Κρητεῖναν πατεῖσθαι* *τραγεῖς* *την τραγεῖς* *τραγεῖς*. The final *τραγεῖς* is probably despised. The final

ετεῖ *φέρει*: for this modest *ετεῖ*, cp. Ph. 195 ετεῖ ετεῖ *ετεῖ* *φέρει*, and n. on *Ant.* 195 ετεῖ ετεῖ *ετεῖ* *φέρει*.—*τεῖται*, *χεῖται* *φέρει*: just, though *φέρει ετεῖ φέρει* seems to occur nowhere else, and *τεῖται φέρει* not only in Lucifer, but in the *EL* 1975.—*τεῖται* *χεῖται* *φέρει* *ετεῖ* *τεῖται* *χεῖται* *φέρει* (*τεῖται* *χεῖται* *φέρει* *ετεῖ*).—*τεῖται* *χεῖται* *φέρει* (*τεῖται* *χεῖται* *φέρει* *ετεῖ*) is probably 'looked with favour'. The alternative is to take *τεῖται* as *τεῖται* *χεῖται* *φέρει* (*τεῖται* *χεῖται* *φέρει* *ετεῖ*) to be despised. Cf. Thuc. i. 11 § 1 (the Greeks at Troy) *πατεῖσθαι...τραγεῖς* *τραγεῖς* *την Κρητεῖναν πατεῖσθαι* *τραγεῖς* *την τραγεῖς* *τραγεῖς*. The final

ετεῖ *φέρει*: cp. O. T. 1959 *τραγεῖς* *τραγεῖς*, and id. 1959.—*πράγη*, *ζεῖται*: cp. 21.

Corpora characteristics

- **5 languages**, 2.3 million tokens and **78,000 entities** linked to Wikidata.

Table 3

Overview of newspaper corpora statistics (hipe-2022 release v2.1). NIL percentages are computed based on linkable entities (i.e., excluding time entities for `hipe2020`).

Dataset	Lang	Fold	Docs	Tokens	Mentions				
					All	Fine	Nested	Noisy	SNIL
<code>hipe2020</code>	<code>de</code>	Train	103	86,446	3,494	3,494	158	-	15.70
		Dev	33	32,672	1,242	1,242	67	-	18.76
		Test	49	30,738	1,147	1,147	73	12.55	17.40
	<code>en</code>	Total	185	149,856	5,883	5,883	298	-	16.66
		Train	-	-	-	-	-	-	-
		Dev	80	29,060	966	-	-	-	44.18
	<code>fr</code>	Test	46	16,635	449	-	-	5.57	40.28
		Total	126	45,695	1,415	-	-	-	42.95
		Train	158	166,218	6,926	6,926	473	-	25.26
	<code>es</code>	Dev	43	37,953	1,729	1,729	91	-	19.81
		Test	43	40,855	1,600	1,600	82	11.25	20.23
		Total	244	245,026	10,255	10,255	646	-	23.55
<code>newseye</code>	<code>de</code>	Total	555	440,577	17,553	16,128	944	-	22.82
		Train	7	374,250	11,381	21	876	-	51.07
		Dev	12	40,046	530	5	22	-	22.08

		Docs	Tokens	Mentions (all)	%NIL	
newseye	newspapers	1907	2,211,449	71,114	30.2	
letemps	class. comm.	297	111,218	7,482	1.4	
topres19th	sonar	140	378,111	75,003	41.92	
<code>sv</code>	<code>fr</code>	Total	140	378,111	75,003	
		Train	21	56,307	2,140	
		Dev	21	6,907	266	
		Dev2	21	6,987	311	
		Test	21	16,163	604	
<code>fr</code>	<code>fr</code>	Total	84	86,364	3,321	
		Train	414	379,481	9,159	
		Dev	51	38,650	869	
<code>fr</code>	<code>fr</code>	Test	51	48,469	1,017	
		Total	516	466,600	11,045	
		Train	414	379,481	9,159	
<code>en</code>	<code>en</code>	Dev	34	11,916	236	
		Test	112	43,263	1,186	
		Total	455	179,156	4,601	
<code>de</code>	<code>de</code>	Total	455	179,156	4,601	
		Train	-	-	-	
		Dev	10	17,477	654	
<code>de</code>	<code>de</code>	Test	10	15,464	471	
		Total	20	32,941	1,125	
		Total	20	32,941	1,125	
Grand Total (newseye)			1,907	2,211,449	71,114	
				27,417	3,224	
					30.23	

Corpora characteristics

- **5 languages**, 2.3 million tokens and **78,000 entities** linked to Wikidata.
- Different entity tag sets and annotation schemes.

Dataset	Coarse tag set	Fine tag set	Nesting	Linking
hipe2020	pers	pers.ind pers.coll pers.ind.articleauthor	yes	yes
letemps	org*	org.adm org.ent org.ent.pressagency	yes	yes
	prod*	prod.media prod.doctr	no	yes
	time*	time.date.abs	no	no
	loc	loc.adm.town loc.adm.reg loc.adm.nat loc.adm.sup loc.phys.geo loc.phys.hydro loc.phys.astro loc.coro loc.fac	yes	yes
		loc.add.phys loc.add.elec loc.unk	yes	yes
			yes	yes
newseye	pers	pers.articleauthor	yes	yes
	org	-	yes	yes
	humanprod	-	yes	yes
	loc	-	no	yes
topres19th	loc	-	no	yes
	building	-	no	yes
	street	-	no	yes
sonar	pers	-	no	yes
	loc	-	no	yes
	org	-	no	yes
ajmc	pers	pers.author pers.editor pers.myth pers.other	yes	yes**
	work	work.primlit work.sclit work.fragm	yes	yes**
	loc	-	yes	yes**
	object	object.manuscr object.museum	yes	no
	date	-	yes	no
	scope	-	yes	no

Corpora characteristics

- **5 languages**, 2.3 million tokens and **78,000 entities** linked to Wikidata.
- Different entity tag sets and annotation schemes.
- Disparities of proportions of noisy mentions (info for hipe2020 and ajmc only, from 10 to 35%).

Corpora characteristics

- **5 languages**, 2.3 million tokens and **78,000 entities** linked to Wikidata.
- Different entity tag sets and annotation schemes.
- Disparities of proportions of noisy mentions (info for hipe2020 and ajmc only, from 10 to 35%).
- Differences of lexical overlap between data folds:
ajmc, letemps and topres19th have a mention overlap which is almost twice that of hipe2020, sonar and newseye.

Dataset	Lang.	% overlap	Folds
ajmc	de	31.43	train+dev vs test
	en	30.50	train+dev vs test
	fr	27.53	train+dev vs test
	Total	29.87	
hipe2020	de	16.22	train+dev vs test
	en	6.22	dev vs test
	fr	19.14	train+dev vs test
	Total	17.12	
letemps	fr	25.70	train+dev vs test
sonar	de	10.13	dev vs test
newseye	fr	14.79	train+dev vs test
	de	20.77	train+dev vs test
	fi	6.63	train+dev vs test
	sv	10.36	train+dev vs test
	Total	16.18	
topres19th	en	32.33	train+dev vs test

Corpora characteristics

- **5 languages**, 2.3 million tokens and **78,000 entities** linked to Wikidata.
- Different entity tag sets and annotation schemes.
- Disparities of proportions of noisy mentions (info for hipe2020 and ajmc only, from 10 to 35%).
- Differences of lexical overlap between data folds: ajmc, letemps and topres19th have a mention overlap which is almost twice that of hipe2020, sonar and newseye.
- Differences in terms of NIL percentages (higher for newspapers than for classical commentaries)

	Docs	Tokens	Mentions (all)	%NIL
newspapers	1907	2,211,449	71,114	30.2
class. comm.	297	111,218	7,482	1.4

HIPE 2022 releases

- HIPE-2022 corpus released as a single package composed of **neatly structured and homogeneously formatted** original datasets.
- Preparation steps:
 - conversion to the HIPE format
 - correction of data inconsistencies
 - metadata consolidation
 - re-annotation of parts of the datasets
 - deletion of rare entity types
 - rearrangement or composition of train and dev splits.

Check out the HIPE-2022 repository:

<https://github.com/hipe-eval/HIPE-2022-data>

HIPE-2022-data

HIPE 2022 shared task is a CLEF 2022 Evaluation Lab on named entity recognition and classification (NERC) and entity linking (EL) in multilingual historical documents.

Following the first CLEF-HIPE-2020 evaluation lab on historical newspapers in three languages, HIPE-2022 is based on diverse datasets and aims at confronting systems with the challenges of dealing with more languages, learning domain-specific entities, and adapting to diverse annotation tag sets. The objective is to gain new insights into the *transferability* of named entity processing approaches across languages, time periods, document types, and annotation tag sets.

Key information

- Primary datasets
- HIPE-2022 Releases
- HIPE-2022 Evaluation
- Acknowledgements
- References

Key information

- Visit the [website](#) for general information on the shared task and registration.
- Read the [Participation Guidelines](#) for detailed information about the tasks, datasets and evaluation.
DOI: [10.5281/zenodo.6045602](https://doi.org/10.5281/zenodo.6045602)
- License: HIPE-2022 data is released under a [CC BY-NC-SA 4.0 License](#)
- Where to find the data:
 - in the [data](#) folder
 - in git [releases](#)
 - on [zenodo](#) DOI: [10.5281/zenodo.6375600](https://doi.org/10.5281/zenodo.6375600)
- Release history:
 - 15.02.2022: release [v1.0](#)
 - 22.03.2022: release [v2.0](#)
 - 15.04.2022: release [v2.1](#)
 - 26.04.2022: commit of all-masked test files for bundle 1 to 4 in data v2.1 ([cf. PR#7](#) and release [v2.1-test_allmasked+sonar_fix](#))
 - 05.05.2022: commit of EL-masked test files for bundle 5 in data v2.1 ([cf. PR#10](#))
 - 13.05.2022: release of test files (except topres19th) used for the evaluation on 13.05.2022 ([cf. PR#11](#) and release [v2.1-test](#))
 - 20.05.2022: release of topres19th test file used for the evaluation on 13.05.2022 ([cf. PR#12](#) and release [v2.1-test-all-unmasked](#))

Evaluation Framework: Tracks and Challenges

To accommodate the HIPE-2022 combinatory (tasks, languages, document types, entity tag sets), the evaluation lab is organized around:

- **Track:** a triple composed of test data [dataset-language-task].
Ex: dataset1-language1-NERCCoarse
dataset1-language2-NERCCoarse
dataset2-language1-EL
etc.
- **Challenge:** a predefined set of tracks
(a kind of tournament composed of tracks).

Evaluation Framework: Tracks and Challenges

Challenges guide participation towards the **development of approaches that work across settings**; we specifically evaluate 3 challenges:

- **Multilingual Newspaper Challenge:**
newspaper document type, minimum 2 languages;
- **Multilingual Classical Commentary Challenge:**
classical commentary document type, minimum 3 languages;
- **Global Adaptation Challenge:**
both document types, minimum 2 languages.

The more (good) tracks a team submits for a given challenge, the more chances they have to win the challenge.

More information in the participation guidelines:

<https://doi.org/10.5281/zenodo.6045662>

February 14, 2022 Technical note Open Access

HIPE 2022 Shared Task Participation Guidelines

• Ehrmann, Maud; • Romanello, Matteo; • Doucet, Antoine; • Clematide, Simon

HIPE 2022 Shared Task Participation Guidelines (Identifying Historical People, Places and other Entities).

HIPE-2022 focuses on named entity processing in historical documents covering the period from the 18th to the 20th century and featuring several languages.

For more information, visit:

- HIPE 2022 website: <https://hipe-eval.github.io/HIPE-2022/>
- HIPE 2022 data repository: <https://github.com/hipe-eval/HIPE-2022-data>
- CLEF 2022 website: <https://clef2022.clef-initiative.eu/>

Preview Page: 1 of 29 Automatic Zoom:

HIPE 2022

Participation Guidelines

Identifying Historical People, Places and other Entities

Shared Task on Named Entity Recognition and Linking in Multilingual Historical Documents

HIPE-2022 Challenges & Tracks				
Tracks (flat list of dataset-task-lang)	Multilingual Newspaper Challenge (MNC)	Multilingual Classical Commentary Challenge (MCC)	Global Adaptation Challenge (GAC)	
1 hipe2020-coarse-en 2 hipe2020-coarse-de 3 hipe2020-coarse-fr 4 newseye-coarse-de 5 newseye-coarse-fi 6 newseye-coarse-fr 7 newseye-coarse-sv 8 letemps-coarse-fr 9 topres19th-coarse-en 10 sonar-coarse-de 11 hipe2020-el-en 12 hipe2020-el-de 13 hipe2020-el-fr 14 newseye-el-de 15 newseye-el-fi 16 newseye-el-fr 17 newseye-el-sv 18 topres19th-el-en 19 sonar-el-de*	NERC-Coarse	Multilingual Newspaper		NERC-Coarse
20 ajmc-coarse-de** 21 ajmc-coarse-en 22 ajmc-coarse-fr 23 ajmc-el-de 24 ajmc-el-en 25 ajmc-el-fr 26 hipe2020-fine-de 27 hipe2020-fine-fr 28 newseye-fine-fi 29 newseye-fine-fr 30 newseye-fine-sv 31 newseye-fine-de 32 letemps-fine-fr 33 ajmc-fine-de 34 ajmc-fine-en 35 ajmc-fine-fr	EL		EL	Global Adaptation Coarse
		NERC-Coarse	Multilingual Classical Commentary	NERC-Coarse
	EL			EL
			NERC-Fine	
				NERC-Fine

Evaluation Framework: System Evaluation

- **Measures:** Macro & **Micro** Precision, Recall and F1 measure.
At entity-level, with different evaluation regimes:

	NERC	EL
Strict	exact mention boundaries	consideration of the top link only (overlapping mention boundaries)
Fuzzy	overlapping boundaries	historical mapping, cut-offs @3 and @5 (overlapping mention boundaries)

- **Challenge evaluation:** Given the tracks submitted for a challenge, submitted systems are rewarded points according to their F1-based rank for each track. Points obtained are summed.

HIPE Scorer: <https://github.com/hipe-eval/HIPE-scorer>,

HIPE Evaluation Toolkit: <https://github.com/hipe-eval/HIPE-2022-eval>

Participation and main observations

- 5 teams submitted a total of 103 system runs.
Submitted runs cover all datasets but not all of the 35 possible tracks.
Most of them focusing on NERC-Coarse.
- All systems use **transformer-based approaches with strong pre-trained models**.
- In transfer learning approaches to historical NER, the **selection of pre-trained models** has a considerable impact on performance.
- Systems demonstrated a good ability to adapt to heterogeneous annotation guidelines.
- EL remains considerably more challenging than NERC.

Observations on challenges

“Thematic aggregations” of system rankings across datasets and languages.

- **Multilingual Newspaper Challenge:**
4 teams; unfortunately not enough track submissions to understand which system works best across settings.
- **Multilingual Classical Commentary Challenge:**
3 teams; strong pre-trained model and fine-tuning performs on par with complex transformer architecture. Context enrichment techniques brings less improvement than for newspapers.
- **Global Adaptation Challenge:**
2 teams. Demonstration that systems can be applied to data originating from various domains.

HIPE 2022 Systems and Results

You're invited to find out more on

Thursday 8.09 morning 10:30-12am!

See also the proceedings: <http://ceur-ws.org/Vol-3180/>

Conclusions (1/2)

- **Transferability test** for NERC and EL approaches on challenging historical material;
- New insights in **domain, document type and language adaptation**;
- Overall, **neural-based systems** with strong resources are capable of dealing with historical, noisy and diverse inputs (but we could do better);
- Entity linking remain challenging;

Conclusions (2/2)

Main contributions of the lab:

- First steps towards understanding how approaches can be optimised to perform well across settings in a cultural heritage context;
- Broad historical and multilingual data set;
- Scorer;
- Step towards efficient semantic indexing of historical material.

Future Directions:

Continuing HIPE Evaluation Lab with further document types, languages and NE related tasks (e.g. relation extraction).

Many thanks to

- CLEF organizers
- HIPE 2022 partnering projects:
AjMC, *impresso*, Living with Machines, NewsEye, SoNAR.
- Special thanks to:
Mariona Coll-Ardanuy (LwM), Ahmed Hamdi (NewsEye) and Clemens Neudecker (SoNAR) for their support regarding data provision.
- Members of the advisory board:
Sally Chambers, Frédéric Kaplan and Clemens Neudecker.
- Participating teams (kudos!)

Thank you for your attention

-  HIPE 2022 website: <https://hipe-eval.github.io/HIPE-2022/>
-  HIPE 2022 data: <https://github.com/hipe-eval/HIPE-2022-data>
-  HIPE scorer: <https://github.com/hipe-eval/HIPE-scorer>
-  HIPE 2022 evaluation toolkit: <https://github.com/hipe-eval/HIPE-2022-eval>
-  HIPE-eval Zenodo: <https://zenodo.org/communities/hipe-eval/>
-  @ImpressoProject #clef2022 #hipe2022

Backup slides

HIPE 2022

Evaluation setting overview

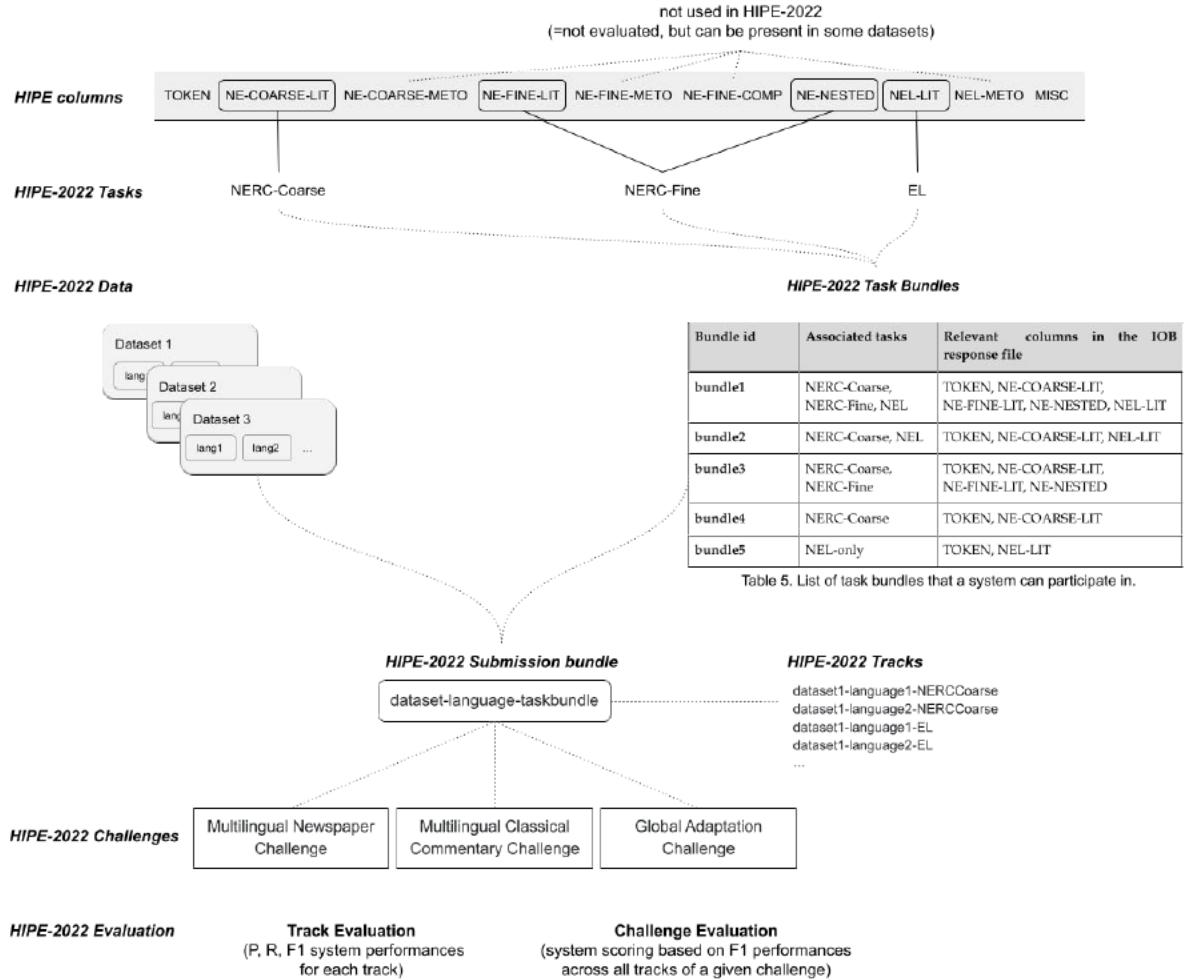


Table 3
Overview of newspaper corpora statistics (hipe-2022 release v2.1). NIL percentages are computed based on linkable entities (i.e., excluding time entities for hipe2020).

Dataset	Lang.	Fold	Docs	Tokens	Mentions			
					All	Fine	Nested	%noisy
nipe2020	de	Train	103	86,446	3,494	3,494	158	- 15.70
		Dev	33	32,672	1,242	1,242	67	- 18.76
		Test	49	30,738	1,147	1,147	73	- 12.55 17.40
	en	Total	185	149,856	5,883	5,883	298	- 16.66
		Train	-	-	-	-	-	-
		Dev	80	29,060	966	-	-	- 44.18
	fr	Test	46	16,635	449	-	-	- 40.28
		Total	126	45,695	1,415	-	-	- 42.95
		Train	158	166,218	6,926	6,926	473	- 25.26
	fr	Dev	43	37,953	1,729	1,729	91	- 19.81
		Test	43	40,855	1,600	1,600	82	- 11.25 20.23
		Total	244	245,026	10,255	10,255	646	- 23.55
Total			555	440,577	17,553	16,138	944	- 22.82
newseye	de	Train	7	374,250	11,381	21	876	- 51.07
		Dev	12	40,046	539	5	27	- 22.08
		Dev2	12	39,450	882	4	64	- 53.74
		Test	13	99,711	2,401	13	89	- 68.52
	fr	Total	44	553,457	15,203	43	1,056	- 49.79
		Train	24	48,223	2,146	15	224	- 40.31
		Dev	24	6,351	223	1	25	- 60.36
		Dev2	21	4,705	203	4	22	- 42.86
		Test	24	14,964	691	7	42	- 47.47
		Total	93	74,243	3,263	27	313	- 41.99
	fr	Train	35	255,138	10,423	99	482	- 42.42
		Dev	35	21,726	752	3	29	- 30.45
		Dev2	35	30,457	1,298	10	63	- 38.91
		Test	35	70,790	2,530	34	131	- 44.82
		Total	140	378,111	15,093	146	705	- 41.92
sv	sv	Train	21	56,307	2,140	16	110	- 32.38
		Dev	21	6,907	266	1	7	- 25.19
		Dev2	21	6,987	311	1	20	- 37.30
		Test	21	16,163	604	0	26	- 35.43
	Total	Total	84	86,364	3,321	18	163	- 32.82
		Train	361	1,092,175	36,790	234	2,237	- 44.36
		Dev	-	-	-	-	-	-
		Dev2	-	-	-	-	-	-
		Test	-	-	-	-	-	-
		Total	516	466,600	11,045	11,045	93	-
letemps	fr	Train	414	379,481	9,159	9,159	69	-
		Dev	51	38,650	869	869	12	-
		Test	51	48,469	1,017	1,017	12	-
	Total	Total	516	466,600	11,045	11,045	93	-
		Train	-	-	-	-	-	-
		Dev	-	-	-	-	-	-
topres19th	en	Train	309	123,977	3,179	-	-	- 18.34
		Dev	34	11,916	236	-	-	- 13.98
		Test	112	43,263	1,186	-	-	- 17.2
	Total	Total	455	179,156	4,601	-	-	- 17.82
		Train	-	-	-	-	-	-
Total	de	Train	-	-	-	-	-	-
	de	Dev	10	17,477	654	-	-	- 22.48
	de	Test	10	15,464	471	-	-	- 33.33
	de	Total	20	32,941	1,125	-	-	- 27.02
Total		Total	20	32,941	1,125	-	-	- 27.02
Grand Total (newspapers)			1,907	2,211,449	71,114	27,417	3,274	30.23

Table 4
Corpus statistics for the ainc dataset (HIPE-2022 release v2.1)

Dataset	Lang.	Fold	Docs	Tokens	Mentions			
					All	Fine	Nested	%noisy
ajmc	de	Train	76	22,694	1,738	1,738	11	13.81
		Dev	14	4,703	403	403	2	11.41
		Test	16	4,846	382	382	0	10.99
	Total		106	32,243	2,523	2,523	13	13.00
en	en	Train	60	30,929	1,823	1,823	4	10.97
		Dev	14	6,507	416	416	0	16.83
		Test	13	6,052	348	348	0	10.34
	Total		87	43,488	2,587	2,587	4	11.83
fr	fr	Train	72	24,670	1,621	1,621	9	30.72
		Dev	17	5,426	391	391	0	36.32
		Test	15	5,391	360	360	0	27.50
	Total		104	35,487	2,372	2,372	9	31.16
Grand Total (ajmc)			297	111,218	7,482	7,482	26	1.45

Table 5
Entity counts by coarse type (HIPE-2022 release v2.1). Although they appear under the same label, identical types present in different data sets may be annotated differently.

	hipe2026			letemps			newseye			sonar			topsres19th			ajmc		
	de	en	fr	fr	de	fi	fr	sv	de	de	en	en	de	en	de	en	f	
pers	1849	558	3706	4086	4061	1212	6201	1132	399				910	844	83			
loc	2923	565	4717	6367	6620	1338	5502	1446	477				3727	43	45	2		
org	652	194	1125	592	3584	350	1758	230	249									
building													563					
street													316					
time	236	46	397															
date														2	20			
prod	223	52	310															
humanprod				6620	1338	5502	1446							12	3	1		
object														465	678	55		
work																		
scope													1091	997	93			

Table 7

Results for NERC-Coarse (micro P, R and F1-score). Bold font indicates the highest, and underlined font the second-highest value.

	Strict			Fuzzy			Strict			Fuzzy			Strict			Fuzzy		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
hipe2020																		
French						German						English						
AAUZH	.718	.675	.696	.825	.776	.800	.716	.735	.725	.812	.833	.822	.538	.490	.513	.726	.661	.692
L3I	.786	.831	.808	.883	.933	.907	.784	.805	.794	.865	.888	.876	.624	.617	.620	.793	.784	.788
NEUR-BSL	.730	.785	<u>.757</u>	.836	.899	<u>.866</u>	.665	.746	.703	.750	.842	.793	.432	.532	.477	.564	.695	.623
letemps																		
French						German						English						
AAUZH	.589	.710	.644	.642	.773	.701	.512	.548	.529	.655	.741	.695	.816	.760	.787	.869	.810	.838
NEUR-BSL	.595	.744	.661	.639	.800	.711	.267	.361	.307	.410	.554	.471	.747	.782	.764	.798	.836	.816
ajmc																		
French						German						English						
HISTeria	.834	.850	.842	.874	.903	.888	.930	.898	.913	.938	.953	.945	.826	.885	.854	.879	.943	.910
L3I	.810	.842	.826	.856	.889	.872	.946	.921	.934	.965	.940	.952	.824	.876	.850	.868	.922	.894
NEUR-BSL	.707	.778	.741	.788	.867	.825	.792	.846	.818	.846	.903	.873	.680	.802	.736	.766	.902	.828
newseye																		
French						German						English						
AAUZH	.655	.657	.656	.785	.787	.786	.395	.421	.408	.480	.512	.495						
NEUR-BSL	.634	.676	.654	.755	.805	.779	.429	.537	.477	.512	.642	.570						
Finnish						Swedish												
AAUZH	.618	.524	.567	.730	.619	.670	.686	.604	.643	.797	.702	.746						
NEUR-BSL	.605	.687	.644	.715	.812	.760	.588	.728	.651	.675	.836	.747						

Table 10

Results for EL-only and End-to-end EL (micro P, R and F1-score @1). For End-to-end EL, only Team L3I submitted runs for hipe-2020. Bold font indicates the highest value.

	Strict			Relaxed			Strict			Relaxed			Strict			Relaxed		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
hipe2020																		
French						German						English						
EL-only																		
L3I	.602	.602	.602	.620	.620	.620	.481	.481	.481	.497	.497	.497	.546	.546	.546	.546	.546	.546
SBB	.707	.515	.596	.730	.532	.616	.603	.435	.506	.626	.452	.525	.503	.323	.393	.503	.323	.393
NIL-BSL	.209	.209	.209	.209	.209	.209	.481	.314	.380	.481	.314	.380	.228	.228	.228	.228	.228	.228
End-to-end EL																		
L3I	.546	.576	.560	.563	.594	.578	.446	.451	.449	.462	.466	.464	.463	.474	.469	.463	.474	.469
sonar																		
German						English												
SBB							.616	.446	.517	.616	.446	.517	.778	.559	.651	.781	.562	.654
Nil-BSL							.333	.333	.333	.333	.333	.333	-	-	-	-	-	-
newseye																		
French						German												
SBB	.534	.361	.431	.539	.364	.435	.522	.387	.444	.535	.396	.455						
Nil-BSL	.448	.448	.448	.448	.448	.448	.485	.485	.485	.485	.485	.485						
ajmc																		
French						German						English						
SBB	.621	.378	.470	.614	.373	.464	.712	.389	.503	.712	.389	.503	.578	.284	.381	.578	.284	.381
Nil-BSL	.037	.037	.037	.037	.037	.037	.049	.049	.049	.049	.049	.049	.046	.046	.046	.046	.046	.046