

Models and Datasets for Cross-Lingual Summarisation

Laura Perez-Beltrachini Mirella Lapata

School of Informatics
University of Edinburgh
`{lperez,mlap}@inf.ed.ac.uk`



Rapid Access to Content Available in a Foreign Language



Variétés locales



'Picholine' et 'Négrette' dans le Languedoc-Roussillon, 'Cailletier' dans les Alpes-Maritimes, 'Salonenque', 'Béruguette' et 'Grossane' dans les Bouches-du-Rhône, 'Bouteillan', 'Cayon' en Provence... Difficile d'établir une liste exhaustive des variétés d'olives qui font la typicité des huit terroirs dotés d'une AOP : Aix-en-Provence, Corse, Haute-Provence, Nice, Nîmes, Nyons, Provence et la vallée des Baux-de-Provence.

Chaque région oléicole se distingue par des variétés prédominantes, d'autres secondaires et les dernières, locales

et anciennes, qui influencent le moment de la récolte, la technique et les conditions d'extraction de l'huile.

Par exemple, Max Doléatto attend **les premières gelées matinales pour la récolte**, «comme mon grand-père, qui était au moulin à Noël». Dans le Gard, en revanche, «*la cueillette commence autour du 1er octobre et s'arrête avant les gelées nocturnes*», explique Jean-René de Fleurieu, propriétaire du château de Montfrin (30).

Récolte et pressage



Les méthodes d'olivaion (récolte des olives) sont variées. Les **machines vibrantes** coexistent avec les méthodes traditionnelles à la main ou **aux peignes**.

Lorsqu'elles ont été gaulées, il faut les récupérer sur des toiles étalées autour des arbres, recueillir dans des filets ou des paniers, et porter les olives au moulin. Là, **elles sont triées** selon leur taille, état et qualité. On les laisse reposer pour qu'elles s'échauffent, sans fermenter. Ensuite, elles sont lavées et broyées avec les noyaux afin d'obtenir une pâte onctueuse.



Cross-Lingual Summarisation Task



Document (French)

Histoire. La consommation alimentaire d'olives sauvages date de la période préhistorique des chasseurs-cueilleurs du Néolithique. L'oléiculture (culture d'oliviers, d'oliveira, et fabrication d'huile d'olive avec des moulins à huile) remonte à la période de l'invention de l'agriculture et de la culture de la vigne et du vin, il y a environ 8 000 ans, dans la région du croissant fertile du Levant au Proche-Orient et en Mésopotamie. L'huile d'olive est alors utilisée pour l'alimentation, la conservation des aliments, la cosmétique, la médecine, les lampes à huile... [...]. Durant la Renaissance du XVe siècle l'Italie devient le plus important producteur réputé d'huile d'olive du monde, avant d'être cultivée à ce jour par l'ensemble des pays du bassin méditerranéen en tant qu'un des fondements de la cuisine méditerranéenne. [...] **Utilisation.** L'huile d'olive est connue depuis la plus haute antiquité : les Grecs anciens, les Phéniciens, les Arabes, les Berbères et les Romains l'utilisaient déjà pour leur cuisine (à l'origine de la cuisine méditerranéenne) et pour leurs produits cosmétiques, ainsi que les Hébreux pour allumer leur chandelle. L'huile d'olive peut être utilisée aussi bien crue (dans des sauces pour salade ou à la place du beurre dans les pâtes par exemple) que cuite (pour la cuisson de viandes ou de légumes ou pour la friture). [...] L'huile d'olive peut également être utilisée pour le traitement du visage, comme le démaquillage des yeux, l'hydratant, l'apaisement des lèvres et la réparation des talons fissurés. Naturellement, l'huile d'olive regorge d'antioxydants anti-âge et de squalène hydratant, ce qui la rend superbe pour les cheveux, la peau et les ongles. Tout comme l'huile de noix de coco, c'est un élément essentiel de tout kit de beauté bricolage. L'huile d'olive est utilisée comme traitement capillaire depuis l'Antiquité égyptienne. [...] **Production.** L'obtention d'un litre d'huile nécessite 4 à 10 kg d'olives suivant la variété d'olive utilisée et son niveau de maturité. La méthode d'extraction utilisée a peu d'incidence. Cependant les moulins utilisant des presses ne peuvent pas utiliser des olives à très forte teneur en eau (à faible rendement en huile) à cause de la fluidité excessive de la pâte. Ceci peut laisser penser, à tort, que leurs rendements sont meilleurs. [...] **Santé.** Par rapport aux autres huiles contenant des acides gras insaturés, l'huile d'olive est assez stable à la cuisson et garde en ce cas ses effets bénéfiques relatifs sur le cholestérol. L'huile d'olive permet de lutter contre le cancer du sein. [...]

Huile d'Olive



Summary (English)

Olive oil is a liquid fat obtained from olives, a traditional tree crop of the Mediterranean Basin, produced by pressing whole olives and extracting the oil. It is commonly used in cooking. It is also used in cosmetics. Olive trees have been grown around the Mediterranean since the 8th millennium BC.

Cross-Lingual Summarisation Task



Document (French)

Histoire. La consommation alimentaire d'olives sauvages date de la période préhistorique des chasseurs-cueilleurs du Néolithique. L'oléiculture (culture d'oliviers, d'oliveira, et fabrication d'huile d'olive avec des moulins à huile) remonte à la période de l'invention de l'agriculture et de la culture de la vigne et du vin, il y a environ 8 000 ans, dans la région du croissant fertile du Levant au Proche-Orient et en Mésopotamie. L'huile d'olive est alors utilisée pour l'alimentation, la conservation des aliments, la cosmétique, la médecine, les lampes à huile... [...]. Durant la Renaissance du XVe siècle l'Italie devient le plus important producteur réputé d'huile d'olive du monde, avant d'être cultivée à ce jour par l'ensemble des pays du bassin méditerranéen en tant qu'un des fondements de la cuisine méditerranéenne. [...] **Utilisation.** L'huile d'olive est connue depuis la plus haute antiquité : les Grecs anciens, les Phéniciens, les Arabes, les Berbères et les Romains l'utilisaient déjà pour leur cuisine (à l'origine de la cuisine méditerranéenne) et pour leurs produits cosmétiques, ainsi que les Hébreux pour allumer leur chandelle. L'huile d'olive peut être utilisée aussi bien crue (dans des sauces pour salade ou à la place du beurre dans les pâtes par exemple) que cuite (pour la cuisson de viandes ou de légumes ou pour la friture). [...] L'huile d'olive peut également être utilisée pour le traitement du visage, comme le démaquillage des yeux, l'hydratant, l'apaisement des lèvres et la réparation des talons fissurés. Naturellement, l'huile d'olive regorge d'antioxydants anti-âge et de squalène hydratant, ce qui la rend superbe pour les cheveux, la peau et les ongles. Tout comme l'huile de noix de coco, c'est un élément essentiel de tout kit de beauté bricolage. L'huile d'olive est utilisée comme traitement capillaire depuis l'Antiquité égyptienne. [...] **Production.** L'obtention d'un litre d'huile nécessite 4 à 10 kg d'olives suivant la variété d'olive utilisée et son niveau de maturité. La méthode d'extraction utilisée a peu d'incidence. Cependant les moulins utilisant des presses ne peuvent pas utiliser des olives à très forte teneur en eau (à faible rendement en huile) à cause de la fluidité excessive de la pâte. Ceci peut laisser penser, à tort, que leurs rendements sont meilleurs. [...] **Santé.** Par rapport aux autres huiles contenant des acides gras insaturés, l'huile d'olive est assez stable à la cuisson et garde en ce cas ses effets bénéfiques relatifs sur le cholestérol. L'huile d'olive permet de lutter contre le cancer du sein. [...]

Huile d'Olive



Summary (English)

Olive oil is a liquid fat obtained from olives, a traditional tree crop of the Mediterranean Basin, produced by pressing whole olives and extracting the oil. It is commonly used in cooking. It is also used in cosmetics. Olive trees have been grown around the Mediterranean since the 8th millennium BC.

- Scarce and costly to create
- Translated and crowd-sourced data
Zhu et al. (2019); Cao et al. (2020); Ouyang et al. (2019); Nguyen and Daumé III (2019)
- WikiLingua how-to guides
Ladhak et al. (2020)

Deriving Cross-Lingual Summarisation Pairs

 **WIKIPÉDIA**
L'encyclopédie libre

cher dans Wikipédia

Article [Discussion](#) Lire [Modifier](#) [Modifier le code](#) [Voir l'historique](#)

Huile d'olive 🌐 84 langues

L'**huile d'olive** est une variété d'**huile alimentaire**, à base de **matière grasse végétale** extraite des olives (fruits d'oliviers cultivés en oliveraie d'oléiculture) lors de la trituration dans un moulin à huile. Elle est un des fondements de la cuisine méditerranéenne (et du paradoxe français) et peut être, sous certaines conditions, bénéfique pour la santé³.

Sommaire [masquer]

Histoire [modifier | modifier le code]


Oliveraie et amphore antique dans la Drôme en Provence.

La consommation alimentaire d'olives sauvages date de la période préhistorique des chasseurs-cueilleurs du Néolithique. L'oléiculture (culture d'oliviers, d'oliveraie, et fabrication d'huile d'olive avec des moulins à huile) remonte à la période de l'invention de l'agriculture et de la culture de la vigne et du vin, il y a environ 8 000 ans, dans la région du croissant fertile du Levant au Proche-Orient et en Mésopotamie. L'huile d'olive est alors utilisée pour l'alimentation, la conservation des aliments, la cosmétique, la médecine, les lampes à huile... Son commerce prend son essor avec le commerce maritime en amphore (surnommée l'or vert de la Méditerranée), vers tout le bassin méditerranéen (histoire du bassin méditerranéen) avec les civilisations antiques phénicienne, grecque, puis romaine (qui améliore et


Répartition de l'oléiculture

Huile d'olive	
	
Indice d'acide	0,3-1,0 ⁴
Indice de saponification	185 - 196 ¹
Propriétés physiques	
T° fusion	-3 °C ¹
Point de trouble	entre 5 et 10 °C.
Point de fumée	180 °C
T° ébullition	300 °C
Masse volumique ρ_{15}^{25}	0,914-0,918 g·cm ⁻³ ²
T° d'auto-inflammation	343 °C
Point d'éclair	225 °C
Viscosité dynamique	à 20 °C : 84 cP (centipoise)
Propriétés optiques	
Indice de réfraction	n_D^{20} 1,466 - 1,468 ¹

Sauper and Barzilay (2009); Zopf (2018); Liu et al. (2018)
Liu and Lapata (2019); Perez-Beltrachini et al. (2019); Hayashi et al. (2021)

Deriving Cross-Lingual Summarisation Pairs

 **WIKIPÉDIA**
L'encyclopédie libre

cher dans Wikipédia

Article [Discussion](#) Lire [Modifier](#) [Modifier le code](#) [Voir l'historique](#)

Huile d'olive 🌐 84 langues

Lead Paragraph

L'**huile d'olive** est une variété d'huile alimentaire, à base de matière grasse végétale extraite des olives (fruits d'oliviers cultivés en oliveraie d'oléiculture) lors de la trituration dans un moulin à huile. Elle est un des fondements de la cuisine méditerranéenne (et du paradoxe français) et peut être, sous certaines conditions, bénéfique pour la santé³.

Sommaire [\(masquer\)](#)

Histoire [\[modifier \]](#) [modifier le code \]](#)


Oliveraie et amphore antique dans la Drôme en Provence.


Répartition de l'oléiculture

La consommation alimentaire d'olives sauvages date de la période préhistorique des chasseurs-cueilleurs du Néolithique. L'oléiculture (culture d'oliviers, d'oliveraie, et fabrication d'huile d'olive avec des moulins à huile) remonte à la période de l'invention de l'agriculture et de la culture de la vigne et du vin, il y a environ 8 000 ans, dans la région du croissant fertile du Levant au Proche-Orient et en Mésopotamie. L'huile d'olive est alors utilisée pour l'alimentation, la conservation des aliments, la cosmétique, la médecine, les lampes à huile... Son commerce prend son essor avec le commerce maritime en amphore (surnommée l'or vert de la Méditerranée), vers tout le bassin méditerranéen (histoire du bassin méditerranéen) avec les civilisations antiques phénicienne, grecque, puis romaine (qui améliore et

Huile d'olive



Indice d'acide 0,3-1,0¹

Indice de saponification 185 - 196¹

Propriétés physiques

T° fusion -3 °C¹

Point de trouble entre 5 et 10 °C.

Point de fumée 180 °C

T° ébullition 300 °C

Masse volumique d_{15}^{15}
0,914-0,918 g·cm⁻³ ²

T° d'auto-inflammation 343 °C

Point d'éclair 225 °C

Viscosité dynamique à 20 °C : 84 cP (centipoise)

Propriétés optiques

Indice de réfraction n_D^{20} 1,466 - 1,468¹

Article Body

Deriving Cross-Lingual Summarisation Pairs

 WIKIPÉDIA L'encyclopédie libre

cher dans Wikipédia

Article Discussion Lire Modifier Modifier le code Voir l'historique

Huile d'olive

84 langues

L'**huile d'olive** est une variété d'**huile alimentaire**, à base de **matière grasse végétale** extraite des olives (fruits d'oliviers cultivés en oliveraie d'oléiculture) lors de la trituration dans un moulin à huile. Elle est un des fondements de la cuisine méditerranéenne (notamment le paradoxe français) et peut être, sous certaines conditions, bénéfique pour la santé³.

 WIKIPÉDIA The free encyclopedia

Article Talk

Not logged in Talk Contributions Create account Log in

Read Edit View history Search Wikipedia

Olive oil

From Wikipedia, the free encyclopedia

Olive oil is a liquid fat obtained from olives (the fruit of *Olea europaea*, family Oleaceae), a traditional tree crop of the Mediterranean Basin, produced by pressing whole olives and extracting the oil. It is commonly used in cooking, for frying foods or as a salad dressing. It is also used in cosmetics, pharmaceuticals, and soaps, and as a fuel for traditional oil lamps, and has additional uses in some religions. The olive is one of three core food plants in Mediterranean cuisine; the other two are wheat and grapes. Olive trees have been grown around the Mediterranean since the 8th millennium BC.

History

Olive oil has long been a common ingredient in Mediterranean cuisine, including ancient Greek and Roman cuisine. Wild olives, which originated in Asia Minor, were collected by Neolithic people as early as the 8th millennium BC.^[1]

Besides food, olive oil has been used for religious rituals, medicines, as a fuel in oil lamps, soap-making, and skin care application.^[citation needed] The Spartans and other Greeks used oil to rub themselves while exercising in the gymnasia. From its beginnings early in the 7th century BC, the cosmetic use of olive oil quickly spread to all of the Hellenic city states, together with athletes training in the nude, and lasted close to a thousand years despite its great expense.^{[17][8]} Olive oil was also popular as a form of birth control; Aristotle in his *History of Animals* recommends applying a mixture of olive oil combined with either oil of cedar, ointment of lead, or ointment of frankincense to the cervix to prevent pregnancy.^[9]

Early cultivation



It is not clear when and where olive trees were first domesticated. The modern olive tree most likely originated in ancient Persia and Mesopotamia, and spread to the Levant and later to North Africa, though some scholars argue for an Egyptian origin.^[10]

The olive tree reached Greece, Carthage and Libya sometime in the 28th century BC, having been spread westward by the Phoenicians.^[10] Until around 1500 BC, eastern coastal areas of the Mediterranean were most heavily cultivated.^[citation needed] Evidence also suggests that olives were being grown in Crete as long ago as 2500 BC. The earliest surviving olive oil

Olive oil



6.6g (constant) (refined and pomace)
0.8% (extra virgin)
184–196 (virgin and refined)
182–193 (pomace)
20 (virgin)
10 (refined and pomace)

This article contains special characters. Without proper rendering support, you may see question marks, boxes, or other symbols.



Oliveraie et amphore antique dans la Drôme en Provence.



Répartition de l'oléiculture

Sommaire

Histoire

La consommation alimentaire d'olives sauvages date de la période préhistorique des chasseurs-cueilleurs du Néolithique. L'oléiculture (culture d'oliviers, d'oliveraie, et fabrication d'huile d'olive avec des moulins à huile) remonte à la période de l'invention de l'agriculture et de la culture de la vigne et du vin, il y a environ 8 000 ans, dans la région du croissant fertile du Levant au Proche-Orient et en Mésopotamie. L'huile d'olive est alors utilisée pour l'alimentation, la conservation des aliments, la cosmétique, la médecine, les lampes à huile... Son commerce prend son essor avec le commerce maritime en amphore (surnommée l'or vert de la Méditerranée), vers tout le bassin méditerranéen (histoire du bassin méditerranéen) avec les civilisations antiques phénicienne, grecque, puis romaine (qui améliore et

Deriving Cross-Lingual Summarisation Pairs

 **WIKIPÉDIA**
L'encyclopédie libre

her dans Wikipédia

Article Discussion Lire Modifier Modifier le code Voir l'historique

Huile d'olive

84 langues

L'**huile d'olive** est une variété d'**huile alimentaire**, à base de **matière grasse végétale extraite des olives (fruits d'oliviers cultivés en oliveraie d'oléiculture)** lors de la trituration dans un moulin à huile. Elle est un des fondements de la **cuisine méditerranéenne** (et du **paradoxe français**) et peut être, sous certaines conditions, **bénéfique pour la santé**³.

Sommaire [masquer]

Histoire [modifier | modifier le code]



Oliveraie et amphore antique dans la Drôme en Provence.



Répartition de l'oléiculture

La consommation alimentaire d'olives sauvages date de la période préhistorique des chasseurs-cueilleurs du Néolithique. L'oléiculture (culture d'oliviers, d'oliveraie, et fabrication d'huile d'olive avec des moulins à huile) remonte à la période de l'invention de l'agriculture et de la culture de la vigne et du vin, il y a environ 8 000 ans, dans la région du croissant fertile du Levant au Proche-Orient et en Mésopotamie. L'huile d'olive est alors utilisée pour l'alimentation, la conservation des aliments, la cosmétique, la médecine, les lampes à huile... Son commerce prend son essor avec le commerce maritime en amphore (surnommée l'or vert de la Méditerranée), vers tout le bassin méditerranéen (histoire du bassin méditerranéen) avec les civilisations antiques phénicienne, grecque, puis romaine (qui améliore et

French Article Body

 **WIKIPÉDIA**
The Free Encyclopedia

Article Talk

Not logged in Talk Contributions Create account Log in

Read Edit View history Search Wikipedia

Olive oil

From Wikipedia, the free encyclopedia

Olive oil is a liquid fat obtained from olives (the fruit of *Olea europaea*, family Oleaceae), a traditional tree crop of the Mediterranean Basin, produced by pressing whole olives and extracting the oil. It is commonly used in cooking, for frying foods or as a salad dressing. It is also used in cosmetics, pharmaceuticals, and soaps, and as a fuel for traditional oil lamps, and has additional uses in some religions. The olive is one of three core food plants in Mediterranean cuisine; the other two are wheat and grapes. Olive trees have been grown around the Mediterranean since the 8th millennium BC.

English

Lead Paragraph

History [edit]

Olive oil has long been a common ingredient in Mediterranean cuisine, including ancient Greek and Roman cuisine. Wild olives, which originated in Asia Minor, were collected by Neolithic people as early as the 8th millennium BC.^[1]

Besides food, olive oil has been used for religious rituals, medicines, as a fuel in oil lamps, soap-making, and skin care application.^[citation needed] The Spartans and other Greeks used oil to rub themselves while exercising in the gymnasia. From its beginnings early in the 7th century BC, the cosmetic use of olive oil quickly spread to all of the Hellenic city states, together with athletes training in the nude, and lasted close to a thousand years despite its great expense.^{[1][2]} Olive oil was also popular as a form of birth control; Aristotle in his *History of Animals* recommends applying a mixture of olive oil combined with either oil of cedar, ointment of lead, or ointment of frankincense to the cervix to prevent pregnancy.^[3]

Early cultivation [edit]



It is not clear when and where olive trees were first domesticated. The modern olive tree most likely originated in ancient Persia and Mesopotamia, and spread to the Levant and later to North Africa, though some scholars argue for an Egyptian origin.^[10]

The olive tree reached Greece, Carthage and Libya sometime in the 28th century BC, having been spread westward by the Phoenicians.^[10] Until around 1500 BC, eastern coastal areas of the Mediterranean were most heavily cultivated.^[citation needed] Evidence also suggests that olives were being grown in Crete as long ago as 2500 BC. The earliest surviving olive oil

Olive oil



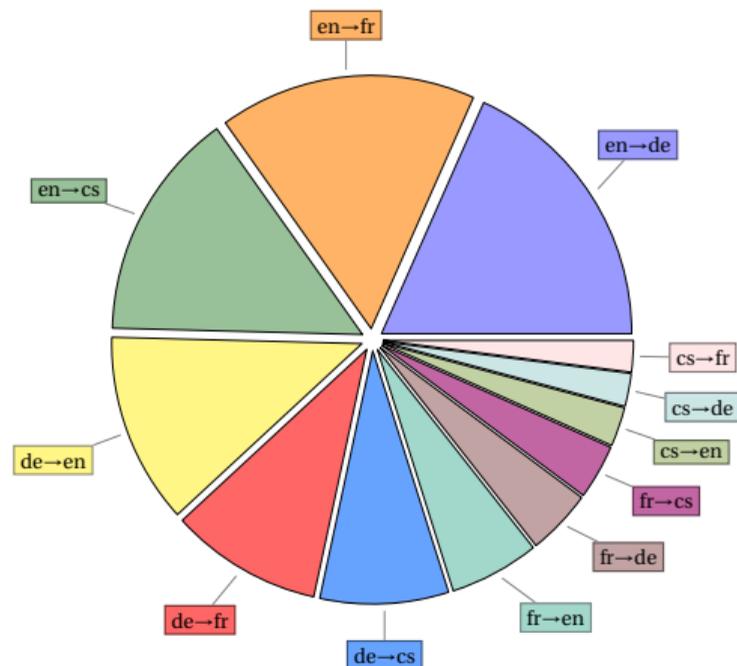
6.6g (monounsaturated)
(refined and pomace)
0.8% (extra virgin)
Saponification value
181–196 (virgin and refined)
182–193 (pomace)
Peroxide value
20 (virgin)
10 (refined and pomace)

This article contains special characters. Without proper rendering support, you may see question marks, boxes, or other symbols.

The XWikis Corpus

English, German, French, Czech

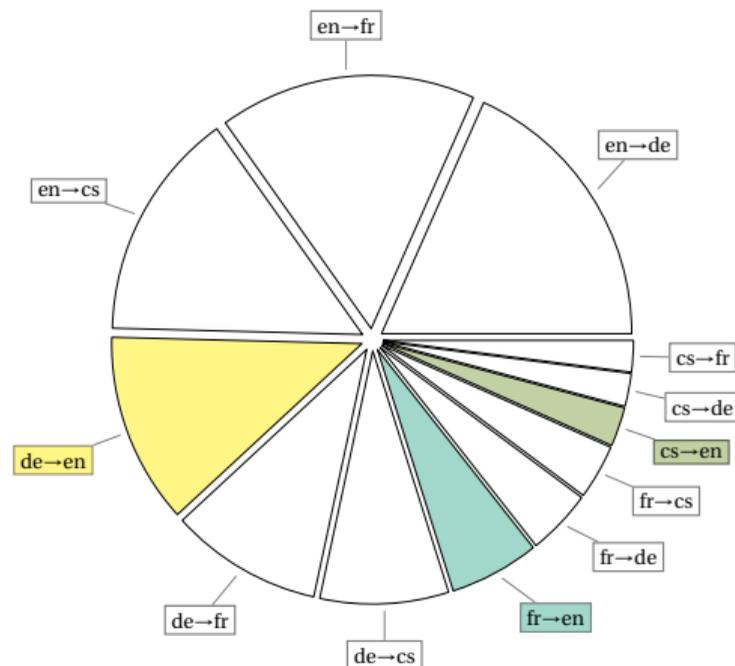
Germanic, Romance, and West Slavic



The XWikis Corpus

English, German, French, Czech

Germanic, Romance, and West Slavic



The summarisation task $X \rightarrow en$

- Human validated overlap 80%
- Documents avg.nb. words: **952**
- Summaries avg.nb. words: **60**
- High compression
- No lead bias
- Poor extractive performance

All-to-English Summarisation $X \rightarrow en$



mBART
mBART50

Liu et al. (2020)

Tang et al. (2020)

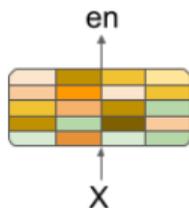
All-to-English Summarisation $X \rightarrow en$



mBART
mBART50

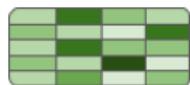
Liu et al. (2020)

Tang et al. (2020)



Supervised
Cross-lingual
Dataset
(Upper-bound)

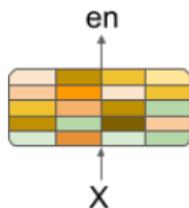
All-to-English Summarisation $X \rightarrow en$



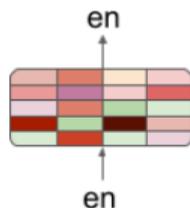
mBART
mBART50

Liu et al. (2020)

Tang et al. (2020)

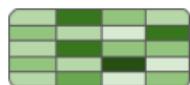


Supervised
Cross-lingual
Dataset
(Upper-bound)



Available
Monolingual
Dataset
(English)

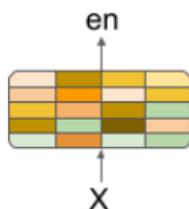
All-to-English Summarisation $X \rightarrow en$



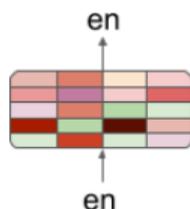
mBART
mBART50

Liu et al. (2020)

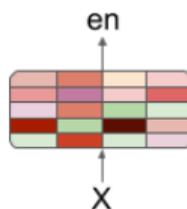
Tang et al. (2020)



Supervised
Cross-lingual
Dataset
(Upper-bound)

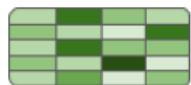


Available
Monolingual
Dataset
(English)



Zero-Shot

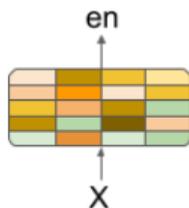
All-to-English Summarisation $X \rightarrow en$



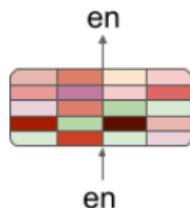
mBART
mBART50

Liu et al. (2020)

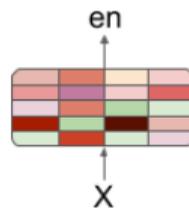
Tang et al. (2020)



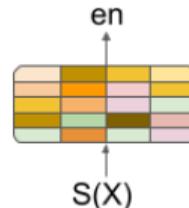
Supervised
Cross-lingual
Dataset
(Upper-bound)



Available
Monolingual
Dataset
(English)

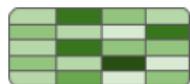


Zero-Shot



Few-Shot
Few
Cross-lingual
Examples

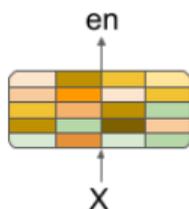
All-to-English Summarisation $X \rightarrow en$



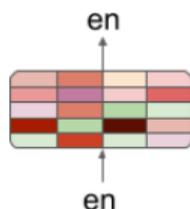
mBART
mBART50

Liu et al. (2020)

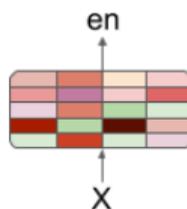
Tang et al. (2020)



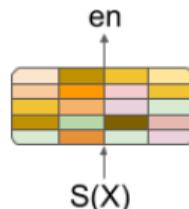
Supervised
Cross-lingual
Dataset
(Upper-bound)



Available
Monolingual
Dataset
(English)



Zero-Shot

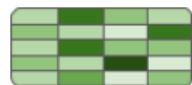


Few-Shot
Few
Cross-lingual
Examples

Zero- and Few-Shot Challenges?

- ▶ understanding input documents in a new language (and new domain)
- ▶ translating relevant content into the target language

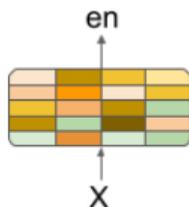
All-to-English Summarisation $X \rightarrow en$



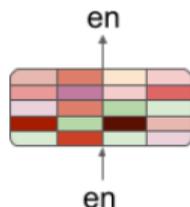
mBART
mBART50

Liu et al. (2020)

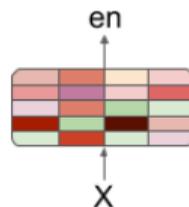
Tang et al. (2020)



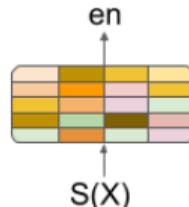
Supervised
Cross-lingual
Dataset
(Upper-bound)



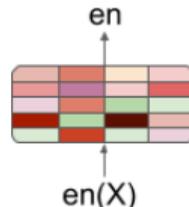
Available
Monolingual
Dataset
(English)



Zero-Shot



Few-Shot
Few
Cross-lingual
Examples



Translated
(Baseline)

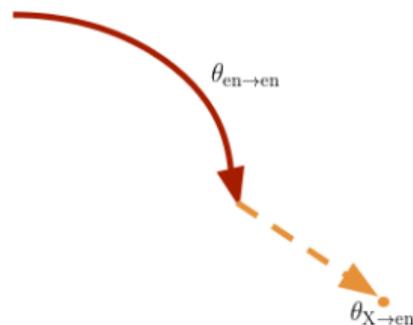
Zero- and Few-Shot Challenges?

- ▶ understanding input documents in a new language (and new domain)
- ▶ translating relevant content into the target language

Few-Shot Approach

- Adapt the English summariser with **few examples** of $X \rightarrow en$
- Reminiscent of the few-shot Model Agnostic Meta-Learning (MAML) algorithm

Finn et al. (2017)

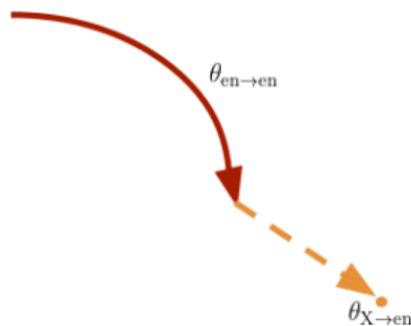


- ▶ Assume the English summariser as meta model
- ▶ Iteration on the cross-lingual task support set $S_{X \rightarrow en}$
- ▶ Meta-update $\theta_{en \rightarrow en}$ w.r.t $\theta_{X \rightarrow en} \approx$ combine $S_{X \rightarrow en}$ with a sample from $en \rightarrow en$
- ▶ Light-weight First Order MAML (**LF-MAML**)

Few-Shot Approach

- Adapt the English summariser with **few examples** of $X \rightarrow en$
- Reminiscent of the few-shot Model Agnostic Meta-Learning (MAML) algorithm

Finn et al. (2017)



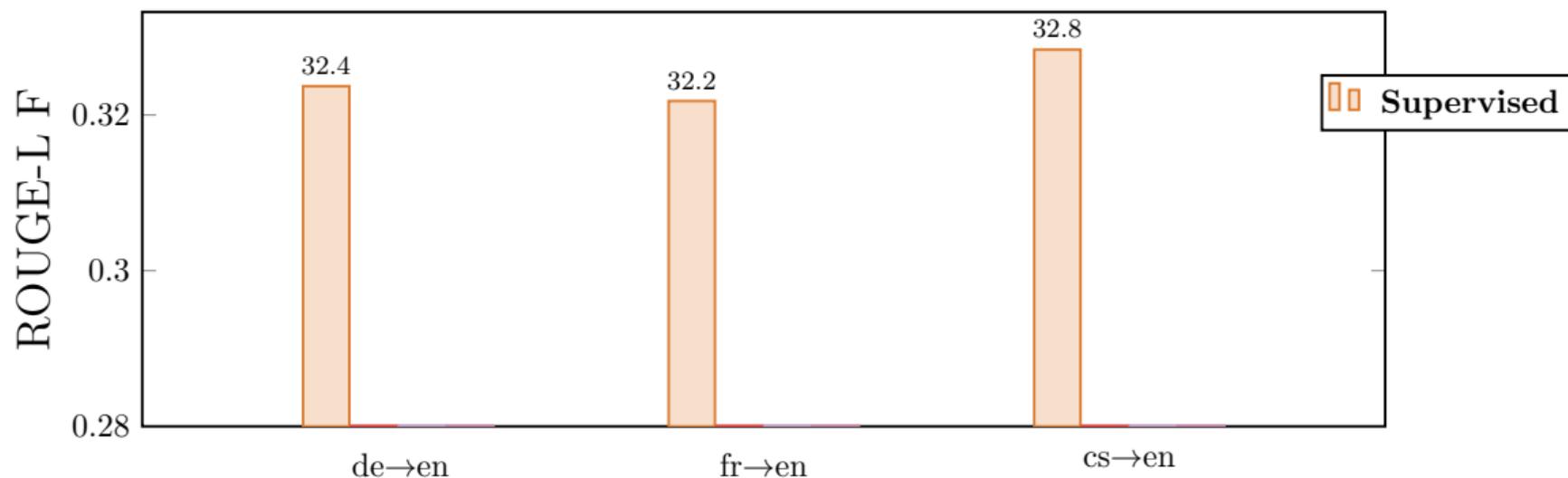
- ▶ Assume the English summariser as meta model
- ▶ Iteration on the cross-lingual task support set $S_{X \rightarrow en}$
- ▶ Meta-update $\theta_{en \rightarrow en}$ w.r.t $\theta_{X \rightarrow en} \approx$ combine $S_{X \rightarrow en}$ with a sample from $en \rightarrow en$
- ▶ Light-weight First Order MAML (**LF-MAML**)

Other variants (in the paper):

- Freezing parameters
- Unlabelled examples (Cross-View Training)

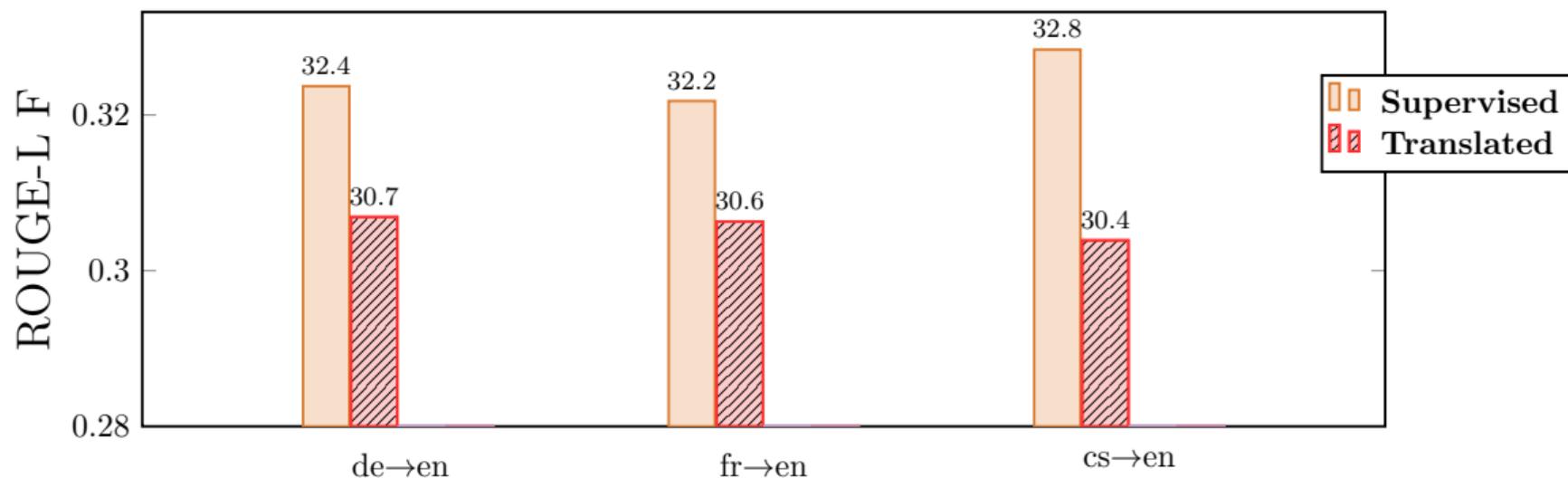
Clark et al. (2018)

Few-Shot Performs Well Across Languages in XWikis



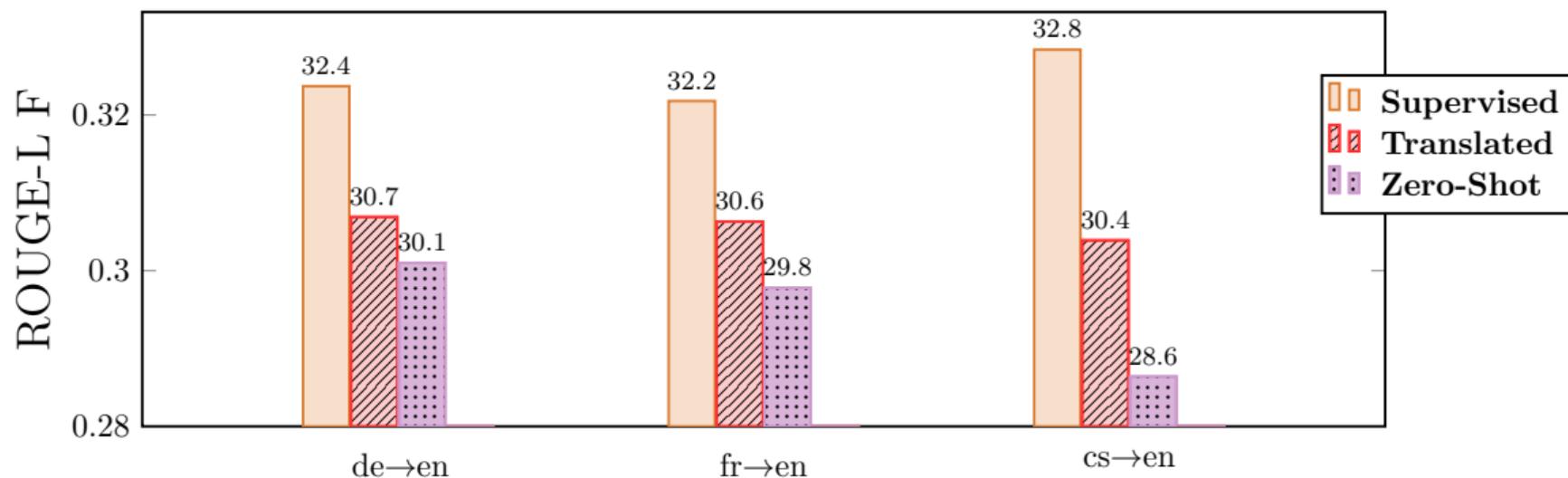
- Similar trend with mBART50; improvements extend to out-of-domain (Voxeurop news)
- 1K LF-MAML (~600 training examples — the rest is used for validation); 7k test examples

Few-Shot Performs Well Across Languages in XWikis



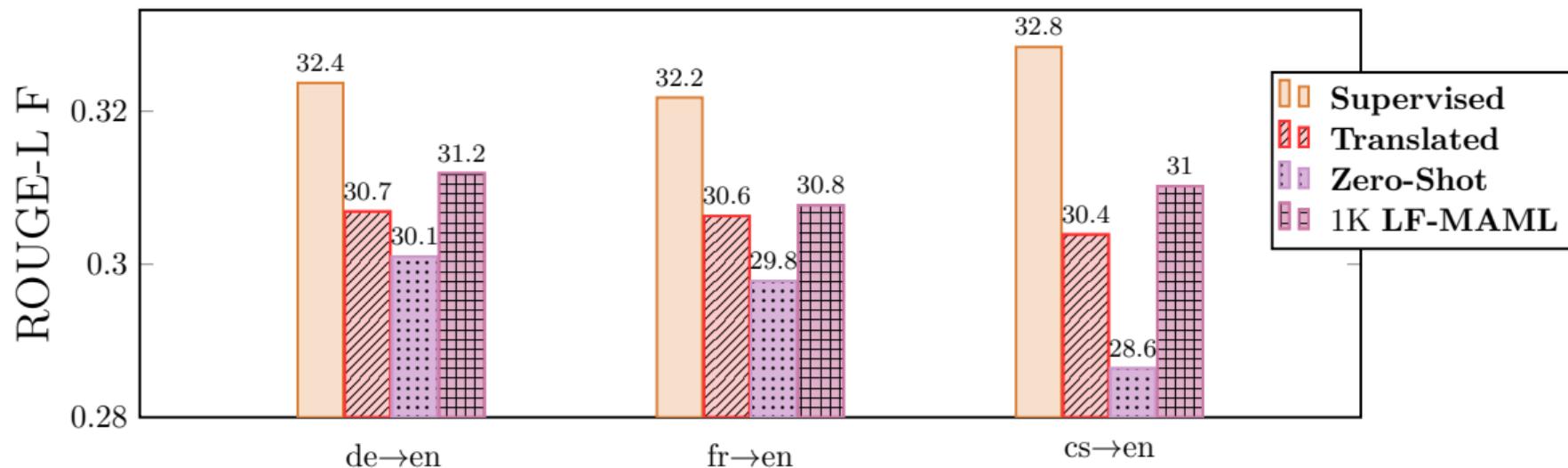
- Similar trend with mBART50; improvements extend to out-of-domain (Voxeurop news)
- 1K LF-MAML (~600 training examples — the rest is used for validation); 7k test examples

Few-Shot Performs Well Across Languages in XWikis



- Similar trend with mBART50; improvements extend to out-of-domain (Voxeurop news)
- 1K LF-MAML (~600 training examples — the rest is used for validation); 7k test examples

Few-Shot Performs Well Across Languages in XWikis



- Similar trend with mBART50; improvements extend to out-of-domain (Voxeurop news)
- 1K LF-MAML (~600 training examples — the rest is used for validation); 7k test examples

Conclusions

- Create XWikis, a new cross-lingual summarisation corpus
 - ▶ four languages: English, German, French, Czech
 - ▶ 12 cross-lingual datasets, different language pairs and directions
 - ▶ methodology to extend to other languages
- First to adapt the few-shot MAML algorithm to cross-lingual summarisation
- Show that with multi-lingual pretrained models, monolingual summarisation data, and few cross-lingual examples it is possible to achieve good cross-lingual summarisation performance

Code and XWikis @ <https://github.com/lauhaide/clads>

References I

- Yue Cao, Hui Liu, and Xiaojun Wan. 2020. Jointly learning to align and summarize for neural cross-lingual summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6220–6231, Online. Association for Computational Linguistics.
- Kevin Clark, Minh-Thang Luong, Christopher D. Manning, and Quoc Le. 2018. Semi-supervised sequence modeling with cross-view training. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1914–1925, Brussels, Belgium. Association for Computational Linguistics.
- Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1126–1135. PMLR.
- Hiroaki Hayashi, Prashant Budania, Peng Wang, Chris Ackerson, Raj Neervannan, and Graham Neubig. 2021. WikiAsp: A Dataset for Multi-domain Aspect-based Summarization. *Transactions of the Association for Computational Linguistics*, 9:211–225.
- Faisal Ladhak, Esin Durmus, Claire Cardie, and Kathleen McKeown. 2020. WikiLingua: A new benchmark dataset for cross-lingual abstractive summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4034–4048, Online. Association for Computational Linguistics.
- Peter Liu, Mohammad Saleh, Etienne Pot, Ben Goodrich, Ryan Sepassi, Lukasz Kaiser, and Noam Shazeer. 2018. Generating Wikipedia by summarizing long sequences. In *Proceedings of the 6th International Conference on Learning Representations*, Vancouver, Canada.
- Yang Liu and Mirella Lapata. 2019. Hierarchical transformers for multi-document summarization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5070–5081, Florence, Italy. Association for Computational Linguistics.
- Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual Denoising Pre-training for Neural Machine Translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.
- Khanh Nguyen and Hal Daumé III. 2019. Global voices: Crossing borders in automatic news summarization. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pages 90–97, Hong Kong, China. Association for Computational Linguistics.
- Jessica Ouyang, Boya Song, and Kathy McKeown. 2019. A robust abstractive system for cross-lingual summarization. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2025–2031, Minneapolis, Minnesota. Association for Computational Linguistics.
- Laura Perez-Beltrachini, Yang Liu, and Mirella Lapata. 2019. Generating summaries with topic templates and structured convolutional decoders. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5107–5116, Florence, Italy. Association for Computational Linguistics.

References II

- Christina Sauper and Regina Barzilay. 2009. Automatically generating Wikipedia articles: A structure-aware approach. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 208–216, Suntec, Singapore. Association for Computational Linguistics.
- Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. 2020. Multilingual translation with extensible multilingual pretraining and finetuning.
- Junnan Zhu, Qian Wang, Yining Wang, Yu Zhou, Jiajun Zhang, Shaonan Wang, and Chengqing Zong. 2019. NCLS: Neural cross-lingual summarization. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3054–3064, Hong Kong, China. Association for Computational Linguistics.
- Markus Zopf. 2018. Auto-hMDS: Automatic construction of a large heterogeneous multilingual multi-document summarization corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).