Methodology in journalistic translation

Lecture 2 · How do we study news translation? Methods and research design



Léa Huotari
lea.Huotari@utu.fi

TURUN YLIOPISTO

Outline

- Translating without a source text?
- Debrief: how far back could you get?
- Why a missing source text is a methodological problem
- A map of methods: product- and process-oriented research
- Reconstructing translation chains: looking for fingerprints
- Collecting news data: practical and ethical issues
- Worked example: sourcing and quoting in Helsingin Sanomat
- Workshop: design a mini-study
- Open questions and take-aways

Where is the source text?

Five properties of news source texts

TURUN YLIOPISTO

Journalistic translation research and source texts

- challenges the traditional binary and unidirectional view of translation (Delabastita 2008: 239) which has been the dominant view on source texts in translation studies (Tymoczko 1999; Delabastita 2008). See also van Doorslaer 2010a; Davier and van Doorslaer 2018)
- has contributed to the broadening of the object in translation studies: “Such a working procedure, where a multiplication of source texts is combined with stages of rewriting and transformation, ‘problematizes the existence and status of the source text as commonly understood in a “normal” translational relationship’ (van Doorslaer 2010: 181).

Translation without a source text? (cf. Davier & van Doorslaer 2018)

- The most visible form of translation and interpreting happens at the stage of editing – or news writing (where ‘writing’ has to be understood in a broad multimodal sense). (Davier 2022: 210)
- In journalistic translation, source texts are
  - multiple
  - fragmented
  - multilayered (may include indirect or intralingual translation)
  - unstable
  - recreated (Cf. Huotari 2026)

Multiplicity of Source Texts

- Journalists draw on multiple texts with multiple authors from multiple sources
- Source texts are present at every stage of the news reporting process:
  - Planning and keeping informed
  - Gathering information
  - Selecting pertinent data
  - Writing and recontextualising information
  - Editing content
- Source texts include:
  - Traditional first hand journalistic sources: interviews, speeches, press conferences
  - second hand sources from other media outlets (news agencies, press releases newspapers, magazines, TV, radio, Social media and the internet
- The final news item typically contains only fragments of several source texts

TURUN YLIOPISTO

Fragmented Source Texts

- Journalists do not translate source texts in their entirety, but rather weave a kind of patchwork from multiple fragments of different source texts (van Doorslaer 2010b)
- Davier and van Doorslaer (2018) advocate for working with short units of translation and not a whole text.

Multilayered or Mediated Source Texts

- “News outlets publish more copy which has been generated by other, external news agencies than by their own in-house journalist. […] “[t]he average international news item published anywhere in the world has probably been through at least four separate newsrooms.” “[…] At each stage, copy is received, put through the newsroom’s editing process and transmitted on to the next receiver, where the cycle is repeated.” (Bell 1991: 47)
  - source texts are usually difficult, if not impossible, to trace a posteriori:
  - “[t]he multiple layers of processing are usually not explicitly evident in the final text, as editors aim to produce a unified text concealing their intervention and manipulation.” Vuorinen (1995: 96)
- This material is often taken for granted by journalists, and thus, rarely checked (Diekerhof & Bakker, 2012: 252).

Unstable source texts

- Dichotomy between stable and unstable source texts (Hernández Guerrero 2009, 43–46)
  - stable sources: authorship is maintained (e.g. opinion articles, columns) translated faithfully
  - unstable sources: the freedom to change, rewrite, update texts
  - “The medieval texts were in fact constantly being rewritten, and translation was frequently perceived as just another step in that chain of rewritings.” (Pym 2004, 175).
  - news presented as translation is extremely rare (from 0-4,6%, in van Doorslaer 2009: 88)

Recreated Source Texts

- News translations “are not straightforward and faithful reproductions of their source texts, as often assumed by lay-people” (Schäffner 2008: 22)
- conceptualization of source text as raw material in the journalistic work
- [t]ranslating and writing are intertwined in one process that is creative and re-creative at the same time. It is impossible to separate the two activities of this integrated process. It is also impossible to distinguish between these two functions in a newsroom: journalists writing “original” reports or journalists translating and rewriting on the basis of existing sources. (van Doorslaer 2010: 181–182)
- It can result in pseudo source texts: a COVID-19 news item presented as a translation (foreign name, source credits) was in fact written by the journalist to “build credibility”, assuming readers could not check the English original (Phanthaphoommee 2023: 29).

To sum up...

- News « is all about retelling » (Catenaccio et al., 2011: 1844), which implies the extraction of text from one discourse (decontextualization) and its reinsertion into another context (recontextualization) (Bauman and Briggs: 1990; Silverstein and Urban,1996; Davier 2022: 401-402)
  - source texts are largely readapted to the journalistic needs of the target language and culture to respond to their new context of use.
- boundaries between the discourse of the journalists and the original writers' tend to be blurred, providing no clear-cut distinction between the two (Kang 2007: 221)

Debrief: how far back could you get?

What made it easy or difficult to trace the Académie Goncourt's or von der Leyen's words? Consider:

Information about sources in the article itself (credit lines, "according to…").

Access to the original: the communiqué (A); the prepared text and the recording of the speech (B).

Language barriers.

The nature of news production (e.g. reliance on agencies, several versions of the same quotation).

Case A · Goncourt

- Did you find the communiqué itself?
- Was « selon toute vraisemblance » kept – and in what form?

Case B · Mercosur

- Prepared text, delivered speech or AFP?
- How was "fair trade" rendered?

TURUN YLIOPISTO

Debrief: one source, many versions

Académie Goncourt (French)	Le monde with AFP (English)*	Reuters (English)**
« selon toute vraisemblance très largement le produit d'une Intelligence Artificielle »	"in all likelihood, very largely the product of artificial intelligence"	— (frames it as "accusations he used AI")
« promouvoir et couronner la littérature écrite par des femmes et des hommes »	"promote and honour literature written by women and men"	"promote and honour literature written by women and men"
« nous ne cautionnons ni n'adoubons [...] la création de textes générés avec l'aide de l'IA »	"we do not endorse or sanction in any way the creation of texts generated with the help of AI"	fully against the "creation of texts generated with the assistance of AI"
« Balance ton Claude » (name of the X account)	"Balance ton Claude" (Expose Your Claude)	"Balance ton Claude" (no gloss a little-known X account)

Sanction: approve or punish? An ambiguous choice for adouber (validate, approve) (literally "to knight").

*https://www.lemonde.fr/en/pixels/article/2026/09/25/french-literary-prize-removes-novel-from-longlist-after-claims-of-using-ai_6757946_13.html?srsltid=AU7gw4VCZlVZHqnktwt3_aUvKJBxPwrcSFjFqlJm5EHlRALN1xNOIE73

** https://www.reuters.com/business/media-telecom/frances-prestigious-goncourt-prize-drops-novel-after-author-accused-using-ai-2026-09-25/

TURUN YLIOPISTO

Case A in our languages

Did the Académie's hedge survive – and what else changed?

Language · outlet	How the Académie (or the accusation) is quoted	What changed
French · AFP Brut, Noovo, Le Matin	*« est selon toute vraisemblance très largement le produit d'une intelligence artificielle »*	Hedge kept. AFP lowercases the communiqué's « Intelligence Artificielle » and adds "one of the season's bestsellers" outside the quotation marks. TelQuel (Morocco) shrinks the hedge to one word: « vraisemblable ».
Spanish · AFP Perfil, El Diario de Hoy	*"con toda probabilidad, en gran medida producto de una inteligencia artificial"*	Hedge kept, but très ("very") is lost; couronner → coronar. Perfil's subheading presents the decision as the result of an anonymous X user's accusation.
Italian · Il Post own report, 25.9 · Domani	*accused of using AI "per scrivere per intero il suo libro"*	The X account's « presque intégralement » loses its "almost". The title is translated (O così o la morte); Domani calls the Goncourt "lo Strega francese".
English · Reuters copy Mudgee Guardian (Australia) · AFP: The Local	*Reuters: "accusations he used AI" · AFP: "in all likelihood, very largely…"*	AFP keeps the hedge; Reuters frames the story as accusations. The Conversation drops "very" ("in all likelihood, largely"); one aggregator states it as a fact.

Inside the quotation marks the hedge usually survives; in headlines and the outlets' own words, it often disappears.

TURUN YLIOPISTO

Debrief: what survived the journey?

On framing

- Accusation or fact in the headline?
- Did “in all likelihood” survive?
- Which angle: AI, plagiarism, race?

On localization

- Is the book title translated? How?
- “Expose Your Claude”: what happened to #BalanceTonPorc?
- What was added for your readers?

On verification

- Who verified what – and how?
- Which facts differ between outlets?
- How do we report allegations against a living author?

On verification: what we found

Who verified what – and how?

- “Balance ton Claude” (X): Pangram detector scores; the account was later claimed by essayist Samuel Fitoussi
- Académie Goncourt: relied on others’ “convergent analyses” – no checks of its own
- AFP: seven other detectors – divergent results
- La Presse, Libération: comparison of texts – documented plagiarism

Plagiarism is shown; AI authorship rests on detector scores that disagree.

Which facts differ?

- Plagiarised texts dated 2012 and 2013 (Reuters, citing La Presse) or 2010 and 2014 (Livres Hebdo)
- Pauwels: a “book” (Reuters) or a “preface” (Libération, via the Guardian)
- “Shortlist” (DT Next, India) – it was the first selection
- Copies sold: about 35,000 (Guardian) – nearly 65,000 (TelQuel, a week later)

Each discrepancy points to a different source.

Reporting allegations against a living author

- Attribute: who says it, on what basis
- Keep the hedges: selon toute vraisemblance, “accused of”
- Separate documented plagiarism from disputed AI use
- Give the denial – he calls the allegation racially motivated
- Never state the allegation as fact in a headline

Dropping “almost”, très or selon toute vraisemblance changes the claim, not the style.

Three English versions: Which source text was used in the outlet you chose?

The speech was filmed – and there are already three versions in English:

The prepared text

"…we choose a productive, long-term partnership, and above all…" (Commission, "check against delivery")

No "over isolation".

The delivered speech

"We choose a productive, long-term partnership over isolation." (Reuters video)

"…a very strong message to the world" (Reuters wire)

She added *"over isolation"* while speaking.

AFP's version

"we chose a productive long-term partnership over isolation"

Past tense – a fingerprint of AFP copy.

Several intralingual translation: which version does your outlet follow – the text, the recording or AFP? **In your language:** how is "fair trade" rendered – does it sound like the Fairtrade label?

TURUN YLIOPISTO

Key observations

- Translated news articles may be manipulated to better suit the journalist’s predetermined storylines or institutional rules and conventions.
- Direct quotes are not always verbatim reproductions of what the newsmaker said.
- A comparison of news articles covering the same event can provide clues that guide the readers in detecting such manipulations.

Why is it difficult to trace the source text?

- Translation research traditionally compares a source text with a target text – both must be identifiable and alignable.
- News production involves multiple authors, multiple source texts and several layers of transformation.
- The process can rarely be reconstructed from the published text alone, so parallel corpora are rare (Exceptions : Le monde diplomatique, Courrier International, Internazionale…)
- Triangulation is therefore needed

A map of methods

Product-oriented

What was published?

- Parallel corpora (rarely possible)
- Comparable corpora across languages or outlets
- Reconstructing a “virtual source text” from fragments
- Short units: quotations, headlines (Davier and van Doorslaer 2018)

Can show: What changed between versions

Cannot show: Who changed it, why, and in which order

Process-oriented

How was it produced?

- Newsroom observation (ethnography)
- Interviews with journalists and editors
- Tracking the writing process (e.g. progression analysis, Perrin)

Can show: Decisions, routines and constraints

Cannot show: Patterns across large amounts of text; access can be difficult

Mixed methods

Both

- A corpus combined with interviews or an ethographic study
- e.g. the Helsingin Sanomat study in this lecture

Can show: Patterns and their explanations (triangulation)

Cannot show: More time; two kinds of data to manage

Collecting news data: practical and ethical issues

Archiving

URL, date and time, screenshot or PDF; the Wayback Machine for earlier versions. Articles change after publication.

Access

Paywalls, deleted articles, agency wires that are not public. Record what you could not see.

Searching

Distinctive phrases in the source language, agency credit lines, searches in several languages.

Copyright

Quote short extracts in your thesis; keep full texts as research data only.

Interviews

Informed consent, secure storage of recordings, a data management plan.

Anonymity

Journalists may speak more freely when anonymised (e.g. “Journalist #1”).

A Finnish case study: The Role of Geographical Context in Journalistic Sourcing and Quoting in the Finnish Media

This research is funded by the ASLA-Fulbright Research Grants for Junior Scholars, sponsored by the Fulbright Finland Foundation and the University of Turku and the Cultural Foundation of Finland

Background



Important cuts in the number of foreign correspondents due to financial constraints



Finnish journalists’ concerns about the narrowing range of their language expertise, which has led to an overrepresentation of Finnish and English sources in the Finnish media (Tuomasjukka 2022)



Growing impact of AI on journalism (Guèvremont & Brin 2024; Roy, Brin, & Gramaccia, 2021) and on translation (Liu & Li 2023)

TURUN YLIOPISTO

Impact

- Growing reliance on material from news agencies' services and from major international news media outlets when reporting on foreign countries
  - News agencies' material rarely checked (Diekerhof & Bakker 2012, 252) and is primarily in English
- Reliance on second-hand source material, often originating from a few concentrated English-speaking media outlets,
  - can compromise the diversity of voices and perspectives in the media (Boumans et al., 2016).
  - and, consequently, sustainable foreign news reporting.

Aim and research questions of the study

- Investigate the Impact of Journalists' Geographical Context on sustainable Foreign News Reporting
- How geographic context affects foreign news coverage in terms of variety and plurality of sources?
    - How it affects the use of first-hand sources?
    - How it affects the use of second-hand sources?



Annastina Heikkilä, YLE. https://share.google/images/WMVnYkqjFbw4V0EaV

How was the study designed?

Research question	Material	Coding	Interviews	Triangulation
How does the journalist's geographical context affect sourcing and quoting?	HS articles on the French presidential elections 2007, (2012), 2017, 2022: recurring, comparable events	First-hand vs second-hand sources; elite, experts, ordinary people	11 semi-structured interviews (30–50 min) with journalists and correspondents	Patterns in the corpus explained by practitioners' accounts

A natural contrast

Helsingin Sanomat closed its Paris bureau in 2010.

Discuss

Why France? Why HS? Why elections? Why add interviews to the corpus?

TURUN YLIOPISTO

288 articles published online by *HS* on French presidential elections

2007: 154

- 150 from France **(97%)**
  - 104 from Paris + 17 other locations

1 Brussels
1 Washington
2 from Helsinki

2017: 93 articles

- 81 from Helsinki **(86%)**
- 14 from France (North of France + Paris)
- 2 Helsinki & France

2022: 41 articles

- 33 from Helsinki **(80,5%)**
- 8 articles from France
  - 7 by a freelancer based in the South of France
  - 1 from Paris by a parachute journalist

TURUN YLIOPISTO

Journalist sourcing: Where does the information come from?

First-hand sources

Eyewitness accounts of an event, providing raw, unfiltered information

- Interviews
- Press conferences
- Official speeches

Second-hand sources

- News agency releases
- Other media outlets (Published / broadcasted sources)
- Social media
- Research institutes

Demonstrates the interdependence of media outlets

Translation layers

TURUN YLIOPISTO



Origin of the first-hand sources in the 2017 presidential elections

- Elite (politicians, officials)
- Experts
- Ordinary people

TURUN YLIOPISTO

Quoting of the elite / politicians (2017)



Elite France	Occurences
Macron	58
Le Pen	49
Fillon	23
Hollande	10
Florian Philippot	7
Jean-Marie Le Pen	5
Mélenchon	5
Carla Bruni-Sarkozy	2
Cazeneuve	2
David Rachline	2
Gilles Lebreton	2
Jacques Chirac	2
Macron's mother	2
Valérie Trierweiler	2
Hamon	1
Julien Dray	1
Total French elite	**173**

Quoting the ordinary people



Name of Ordinary People	Occurence
Kati Kukkasniemi	9
Lilie Bultez	9
Lex Paulson	8
Antoine Sauvage	7
Sandra Leman	7
Étienne Cardiles	5
Fouzia	5
Jessica Elessa	5
Jean Philippe Langlierille	4
Jean-Louis Warin	4
Laurence Noret	4
Marie Lepoivre	4
Patrick Houyoux	4
Yohan Chovau	4
Akim	3
Christian Leclaire	3
Dimitry Larivière	3
Dorothée Fizazi	3
Mimoun Lazaria	3
Nicole Geldhof	3
Paul Buigues	3
Sebastien Billet	3
Alexandra Jeannoy	2
Cyril Macke	2
Denis Chassouant	2
Fanny Zafiri	2
Julien Chenen	2
Lewis Lancet	2
Luisa Makhloufi	2
Marion Lefeure	2
Nicole Lancet	2
Terrorists' neighbours	2
Viktor Albrecht	2
Anonym source	1
Maria Mino	1
Total: 35 persons	**127**

Quoting the experts



Greater reliance on a ”network of more or less trusted local contacts” who the journalists already know and ”who are easily and quickly available” (Phillips 2010, 92)

Finnish Expert	Occurrences
Taneli Lahti	19
Aunesluoma	15
Juha Jokela	7
Erka Koivunen, F-Secure	5
Aki Kangasharju	3
Andrew Weisburd	2
Hanna Freystätter	2
Tiina Héléniums	2
Total Finnish	**55**
French Expert	
Christophe Noël	6
Martin Quencez	3
Jean-François Julliard	2
Gilles Ivaldi	1
Mathieu Gallard	1
Total French	**13**
Total Experts	**68**

How journalists see it?

“if you are doing a story here, from Helsinki, then it is easy to focus on things such as power holders or institutions you know, you could quote Reuters, that Macron said this, but then if you are able to go on the place, you can talk to all kinds of people more easily, and yes, and then also, if you are reporting remotely, what is somehow transmitted here through news agencies or, for example, through the French news sites is often a kind of like a kind of dramatic news event, but then the everyday life, the ordinary life or the everyday reality in the background there is never conveyed to the readers.”

Interviewed journalist 5

“The big difference is that, if you're based in Finland, you're translating material that has already been used. You haven't met the people or experienced the situation first-hand. Foreign news reporting is much more than just interviews. For example, if I'm covering the Yellow Vest protests, I can report on the situation as I see it, as well as what I hear people talking about in cafés and what I see on television. Of course, the interviews shown on television news are just the tip of the iceberg, but the most concrete difference is that you're there, so your sources are your own knowledge and first-hand experience. The yellow vests are a good example of how the reality is somehow distorted if you only report it from Finland through television. Certain images have been selected for the news clips distributed by international photo agencies, so you don't get a complete picture of the atmosphere in France or what is happening there. You may also not be able to put it into perspective in the same way. You might think that the whole of Paris is burning.” Journalist#1

Concluding remarks

- The geographic context affects the use of sources
- Reporting from France by a parachute journalist:
  - Ordinary people and sometimes French experts serve as primary sources of information
- Reporting from Helsinki by a local desk journalist:
  - Elite serve as relayed primary sources of information
  - Media sources are varied and diverse but dependant on the language skills of the journalists
  - Primary sources are relayed, sometimes through two or even three media sources.

Open questions

English as a hub language

How many of our pooled versions went through English?

Reliance on a few English-language sources can narrow the range of voices (Boumans et al. 2016).

Machine translation in the newsroom

When journalists use MT, who is the translator?

Does it leave traces in the published text?

AI and pseudo-sources

During the 2026 World Cup, invented quotes and AI-generated videos circulated widely.

One viral caption put words in Haaland’s mouth that his recorded interview contradicts (Euronews, 20.7.2026).

TURUN YLIOPISTO

Bibliography (1/2)

Baker, M. 2006. Translation and Conflict: A Narrative Account. London: Routledge.

Bauman, R., & Charles L. Briggs. 1990. “Poetics and Performance as Critical Perspectives on Language and Social Life.” Annual Review of Anthropology 19: 59–88.

Bell, A. 1991. The Language of News Media. Oxford: Blackwell.

Bielsa, E., & S. Bassnett. 2009. Translation in Global News. London: Routledge.

Davier, Lucile, and Luc van Doorslaer. 2018. “Translation without a Source Text: Methodological Issues in News Translation.” Across Languages and Cultures 19 (2):

241–258.

Diekerhof E. & P. Bakker. 2012. “To check or not to check: An exploratory study on source checking by Dutch journalists." *Journal of Applied Journalism & Media Studies*, 1(2), 241–253.

Doorslaer van, Luc. 2010. “Journalism and Translation.” In Handbook of Translation Studies vol. 1, ed. by Yves Gambier, and Luc van Doorslaer, 180–184. Amsterdam: John Benjamins.

TURUN YLIOPISTO

Bibliography (2/2)

Guèvremont, V., & C. Brin. (Eds.). (2024). Intelligence artificielle, culture et médias. Presses de l'Université Laval. https://doi.org/10.1515/9782763758787.

Huotari, L. (2026). A creative and recreative conceptualization of source text: The case of journalistic translation. In Carlström, M. & R. Pleijel (eds.). A (Re)turn to the Source Text, pp. 88–115.

Kang, J-H. 2007. "Recontextualization of News Discourse: A Case Study of Translation of News Discourse on North Korea." The Translator 13 (2): 219–242.

Liu, X., & Li, C. (2023). Artificial intelligence and translation. In C. Sin-wai (Ed.), Routledge encyclopedia of translation technology, 2nd ed. pp. 280–302). Routledge. https://doi.org/10.4324/9781003168348-16.

Orengo, A. 2005. "Localising News: Translation and the 'Global-National' Dichotomy." Language and Intercultural Communication 5 (2): 168–187.

Roy, J.-H., Brin, C., & Gramaccia, J.A. (Eds.). (2021). Intelligence artificielle: Coup de grâce ou coup de pouce pour le journalisme? [Special dossier]. Cahiers du journalisme et de l'information, NS7. https://cahiersdujournalisme.org/Serie-2/V2N7/.

Sarangi, S. 1998. "Rethinking Recontextualization in Professional Discourse Studies: An Epilogue." Text 18 (2): 301–318.

Valdeón, R. A. 2014. "From Adaptation to Appropriation: Framing the World Through News Translation." Linguaculture 5 (1). https://doi.org/10.1515/lincu-2015-0019.

TURUN YLIOPISTO

Thank you

TURUN YLIOPISTO