



La citation de données de recherche dans des publications en linguistique

– l’histoire d’une initiative prise par
la communauté scientifique

Helene N. Andreassen, PhD

UiT Université arctique de Norvège

Journée évaluation des corpus : où en est-on, où allons-nous ?

Séminaire organisé par CORLI, 3 octobre 2019, Paris

Cette présentation illustre un travail fait en collaboration avec Andrea Berez-Kroeker (U Hawai'i at Manoa), Lauren Gawne (La Trobe U), Lauren Collister (U Pittsburgh), Christopher Cox (Carleton U), Philipp Konzett (UiT), Koenraad De Smedt (U Bergen) & Bradley McDonnell (U Hawai'i at Manoa)



La répliquabilité de la recherche

- *Répliquabilité*: La réutilisation d'une méthode scientifique afin de collecter de nouvelles données qui pourront confirmer des assertions scientifiques.
 - Si la répliquabilité est impossible, comment tester la validité de la recherche?
 - L'accès aux données originales autorise une ré-analyse indépendante.

Si des recherches en linguistique ne sont pas directement répliquables, étant donné l'aspect individuel et contextuel du moment d'enregistrement, un moyen pour valider la recherche est de la reproduire.

La reproductibilité de la recherche

1. Transparence de la méthode appliquée lors de la collecte, du traitement et de l'analyse des données.
2. Transparence du statut des données, à savoir le type (ou la source) des données, où elles se trouvent, et comment on peut y avoir accès.

« I have come to think of the **accuracy of data as a serious problem** for all linguistics journals, and for the field at large. I do not see any way to deal with it unless we constantly remind ourselves and our students of the importance of working to maintain the quality of the data that we use.

[...]

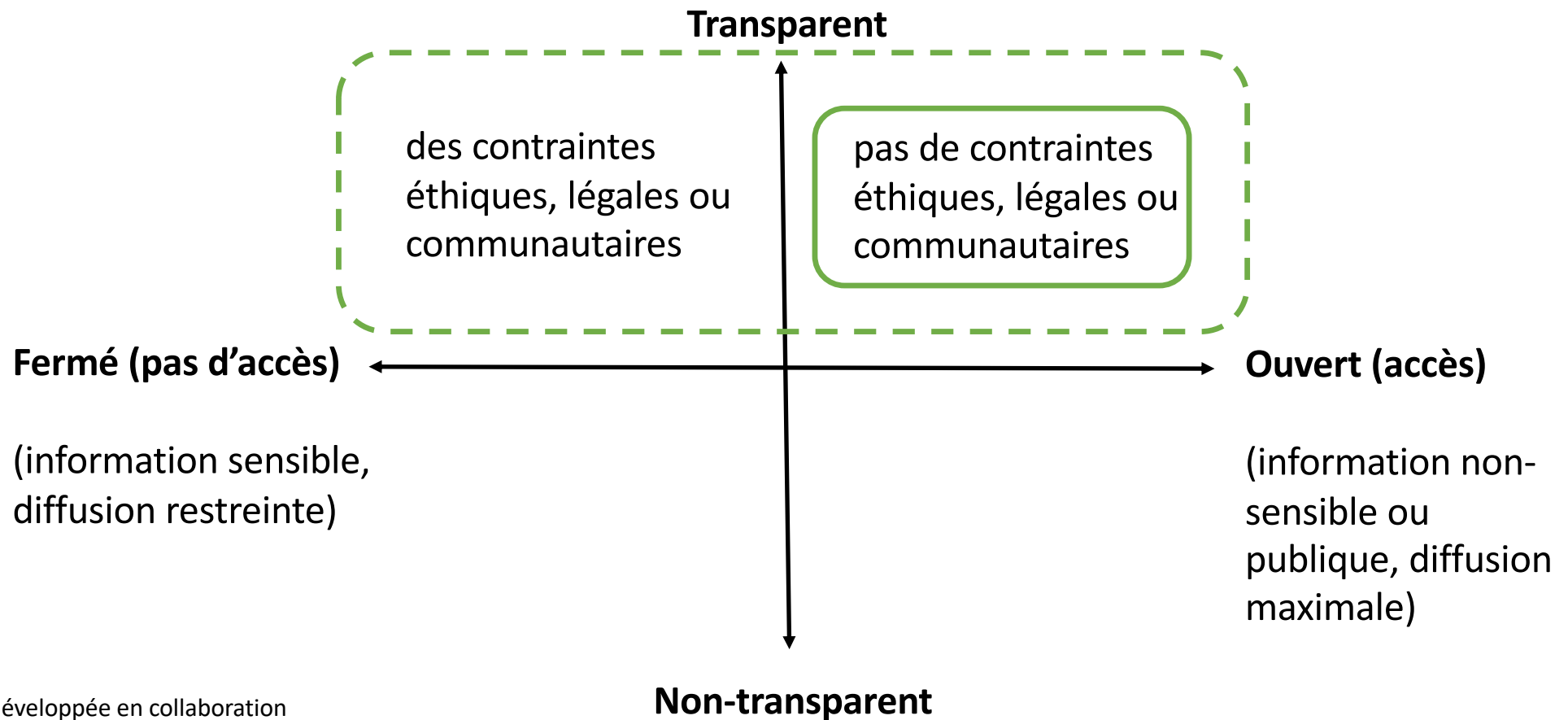
Besides actual errors in cited data, the attribution of data and even well-known analyses to the last linguist who discussed the material in print, rather than to the original published source of the data and/or the analysis, is **so common that it has become a cliché**. »

(Thomason, 1994, p. 410, bolding ours)

Plan de la communication

- La situation idéale et les pratiques réelles
- Une initiative prise par la communauté scientifique
 - L'historique
 - Les objectifs
 - Les résultats
 - Les perspectives
- Défis et réflexion sur l'avenir

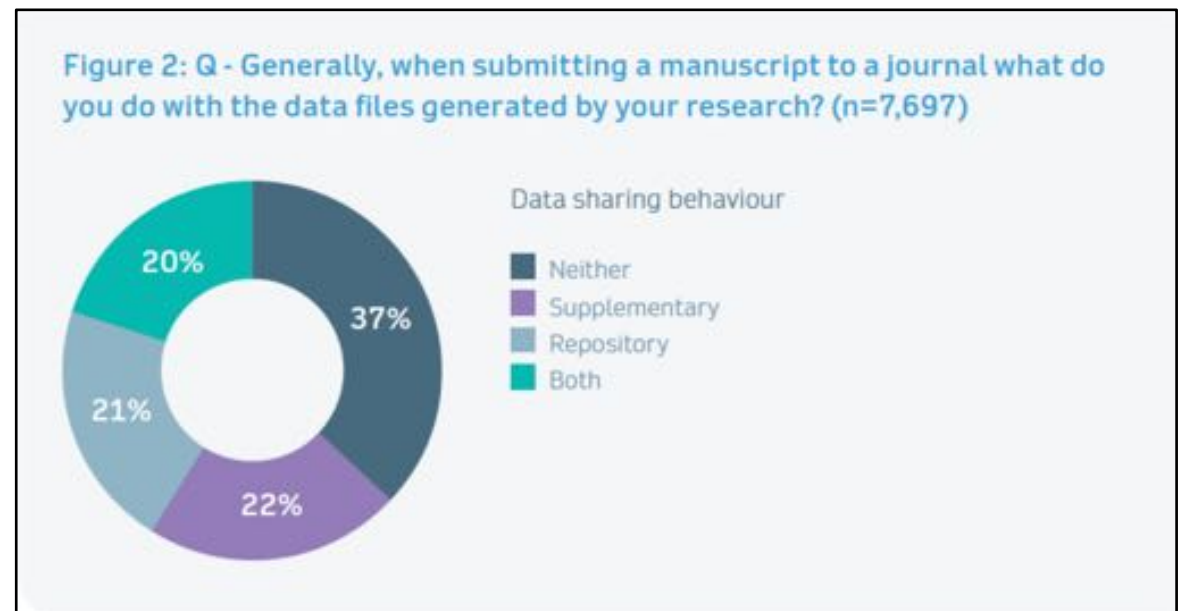
L'idéal: Transparence et libre accès à la recherche



(Figure développée en collaboration avec Aysa Ekanger, UiT)

Pratiques réelles

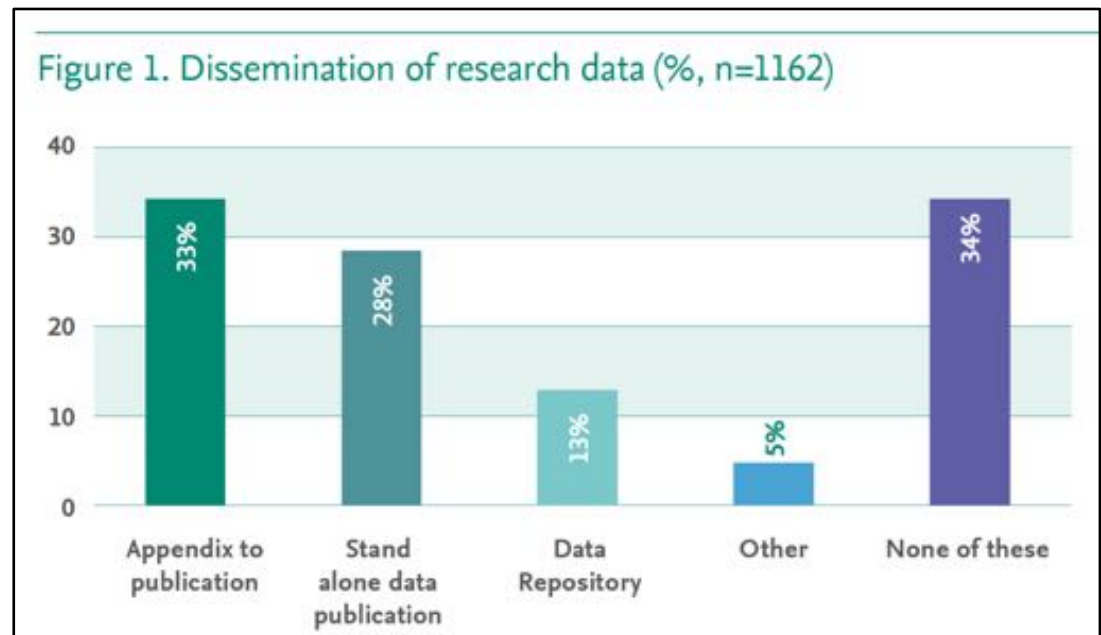
- Plus de 30% des chercheurs ne publient pas leurs données.



Stuart et al, 2018 (N = 7700)

Pratiques réelles

- Moins de 30% des chercheurs publient les données dans une archive.



Wouters & Haak, 2017 (N = 1162)

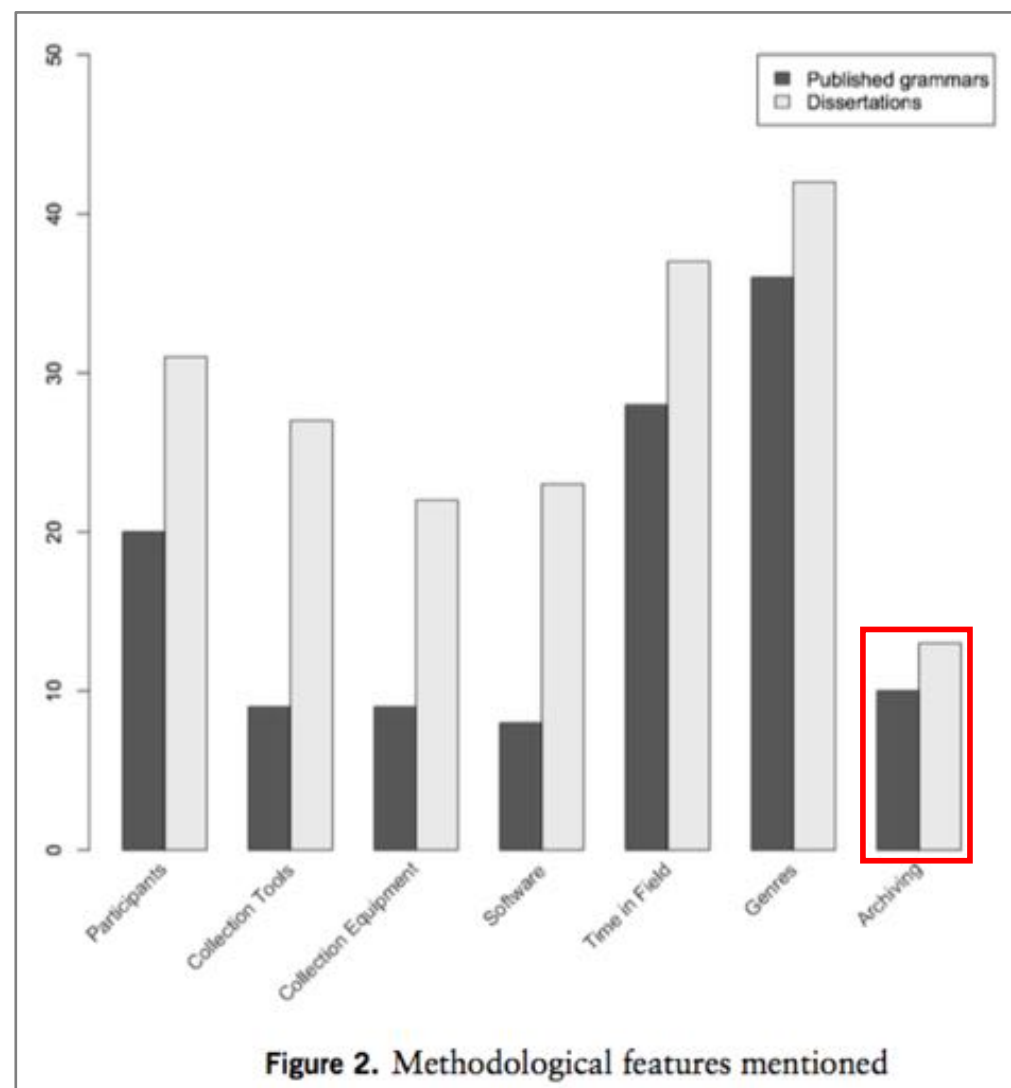
Pratiques réelles en linguistique

Examen de 50 grammaires descriptives et 50 thèses publiées dans la période 2003-2012.

1. Transparence de la méthode

- Informations souvent données sur le type de données, la période de collecte, les participants.
- Pour les grammaires: peu d'informations données sur les outils et l'équipement de collecte, les logiciels.
- Pour les deux types: Peu d'informations sur l'archivage.

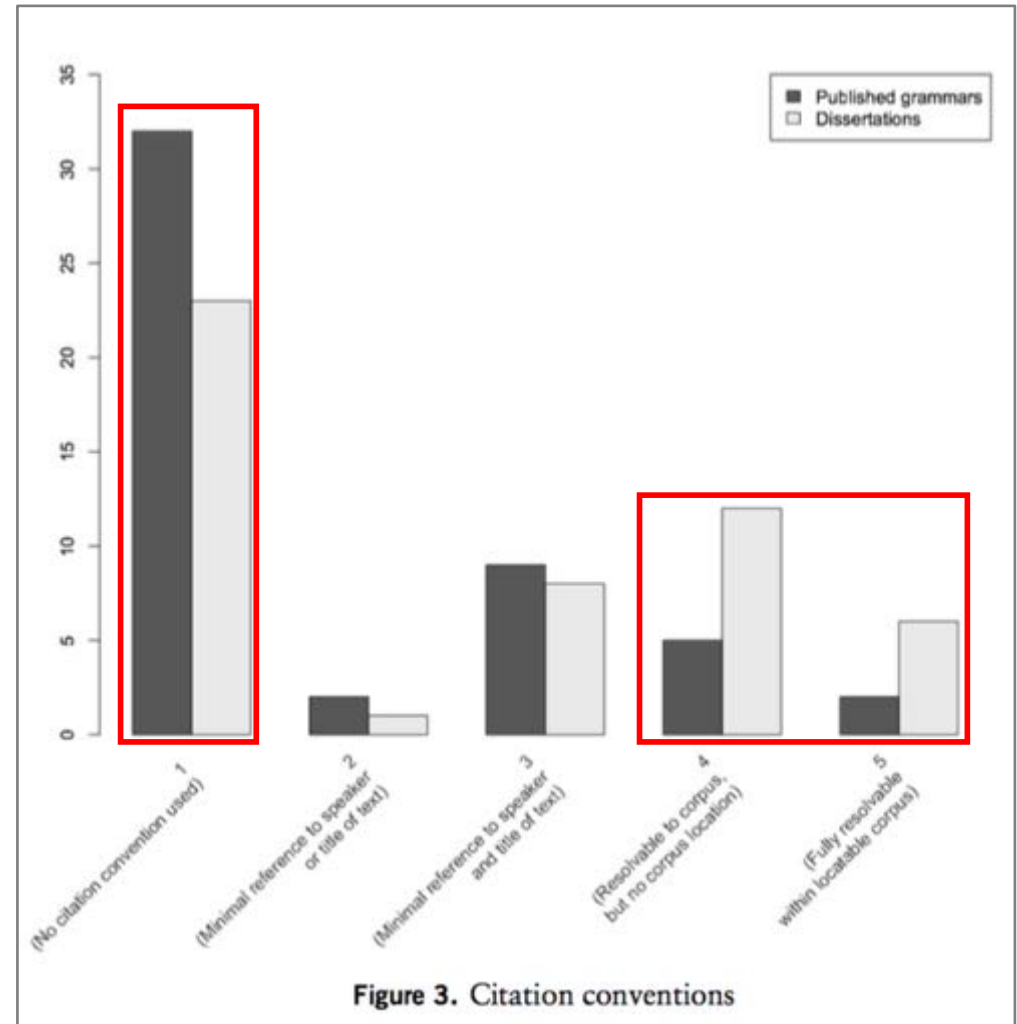
Gawne et al., 2017, p. 172



Pratiques réelles en linguistique

2. Les conventions de citation

- La majorité des publications ne donnent pas de citation à côté des exemples numérotés.
- Observation positive: Un certain nombre de thèses réfèrent à un corpus (avec ou sans indication de sa localisation).

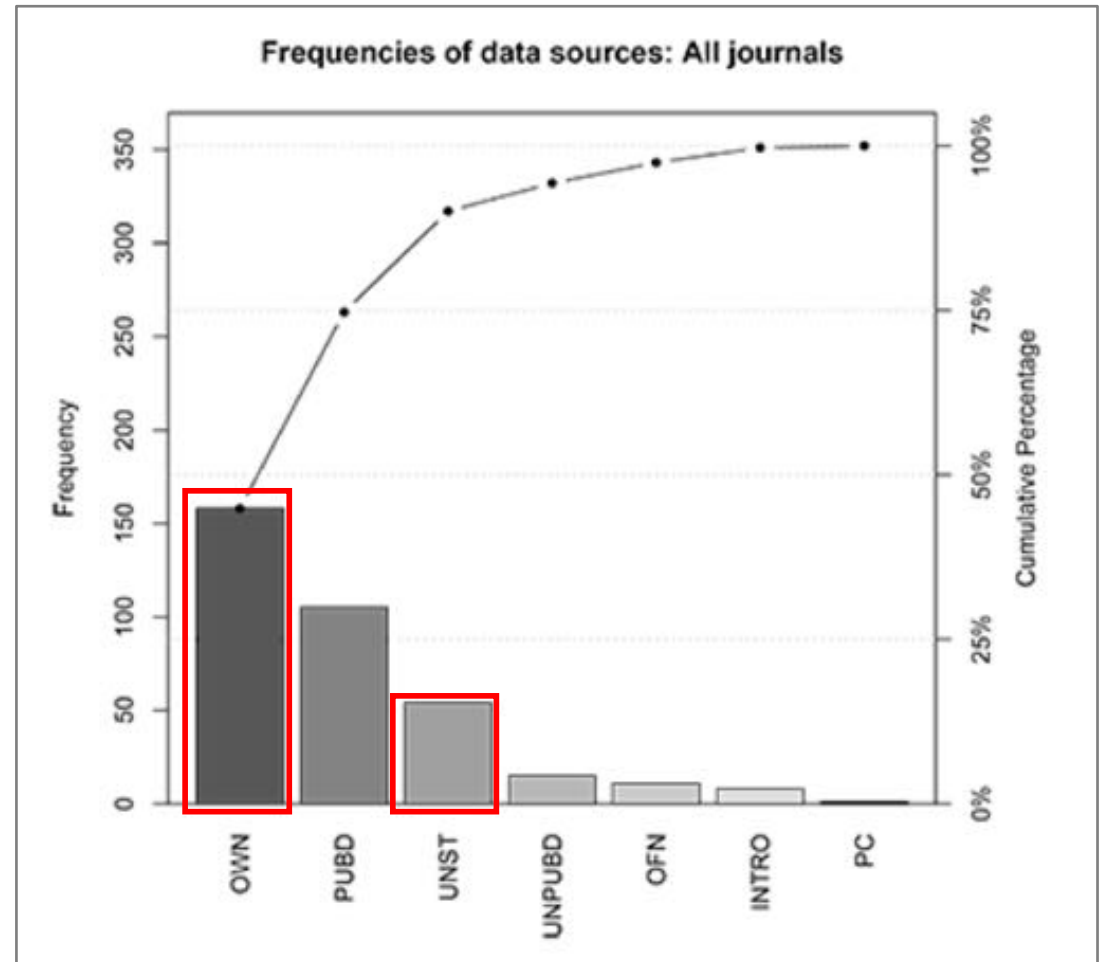


Pratiques réelles en linguistique

Examen des articles de 9 revues en linguistique, publiés dans la période 2003-2012.

1. Transparence du type de données

- Préférence pour l'usage de données collectées par l'auteur lui-même (« OWN »).
- Un certain nombre de publications ne mentionnent pas la source (« UNST »).



Pratiques réelles en linguistique

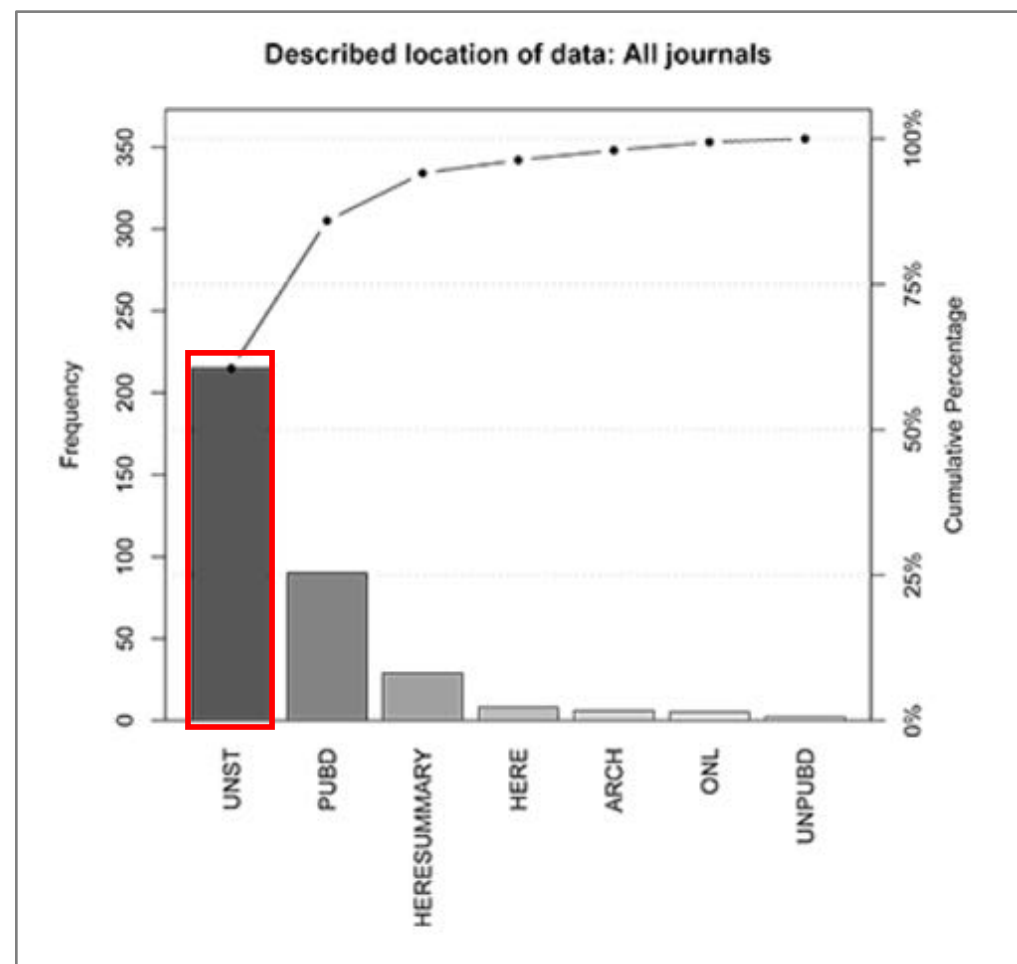
2. Transparence de la localisation

- Plus de la moitié des publications n'explicitent pas la localisation des données.

3. Observation positive

Les différentes sous-disciplines favorisent la transparence pour des éléments concrets.

- Linguistes de terrain: Locuteurs, conditions d'enregistrement.
- Phonéticiens: Equipement de collecte et d'analyse.
- Acquistionnistes: Méthode.



Berez-Kroecker et al., 2017b

La technologie et l'infrastructure
sont en place

Comment améliorer la situation?

Une des initiatives

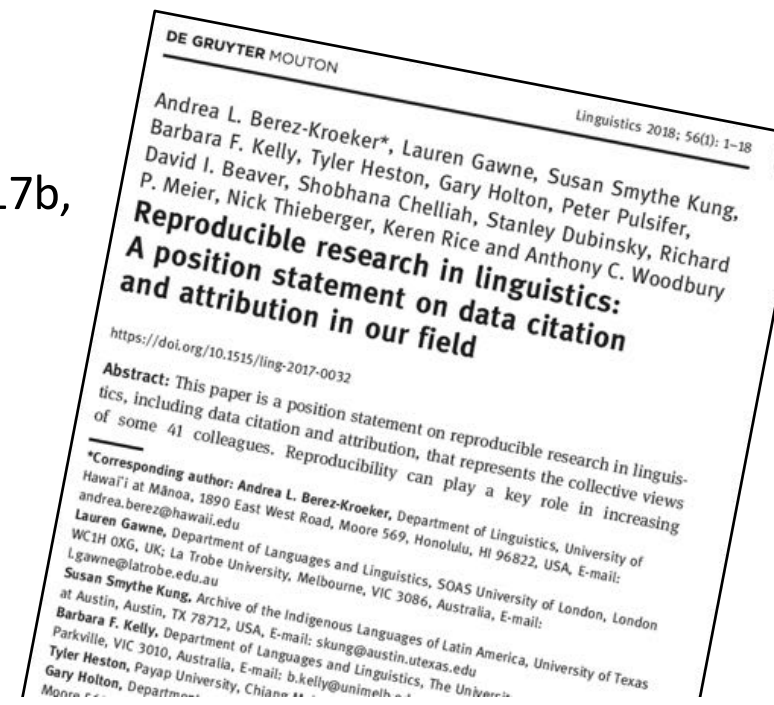
Projet NSF* *Developing Standards for Data Citation and Attribution in Linguistics*

- Débuté en 2015, plus de 40 participants internationaux, 3 workshops, session de panel lors de la rencontre annuelle du LSA (Linguistic Society of America) en 2017.

Résultats

- 2 enquêtes grammairales et revues (Berez-Kroeker et al., 2017b, Gawne et al., 2017).
- Papier de position sur la citation de données linguistiques (Berez-Kroeker et al., 2018).
- Création de LDIG.

*National Science Foundation



Linguistics Data Interest Group (LDIG)

- Réseau international d'environ 100 membres au sein du Research Data Alliance (RDA).
- RDA: Organisation qui vise à créer une infrastructure sociale et technique afin de faciliter le partage libre de données scientifiques.
- Composé de *interest groups*, *working groups* et de groupes nationaux.
- Autour de 9000 membres venant de 137 pays (août 2019), rencontres 2 fois par an.



<https://www.rd-alliance.org/>

Linguistics Data Interest Group (LDIG)

- Co-direction
 - Helene N. Andreassen (UiT)
 - Andrea Berez-Kroeker (U Hawai'i at Manoa)
 - Lauren Gawne (La Trobe U)
- Objectifs principaux
 - Développement et adoption de principes et de guidelines pour la citation et l'attribution des données.
 - Education et sensibilisation.
 - Valorisation du traitement des données dans la profession.

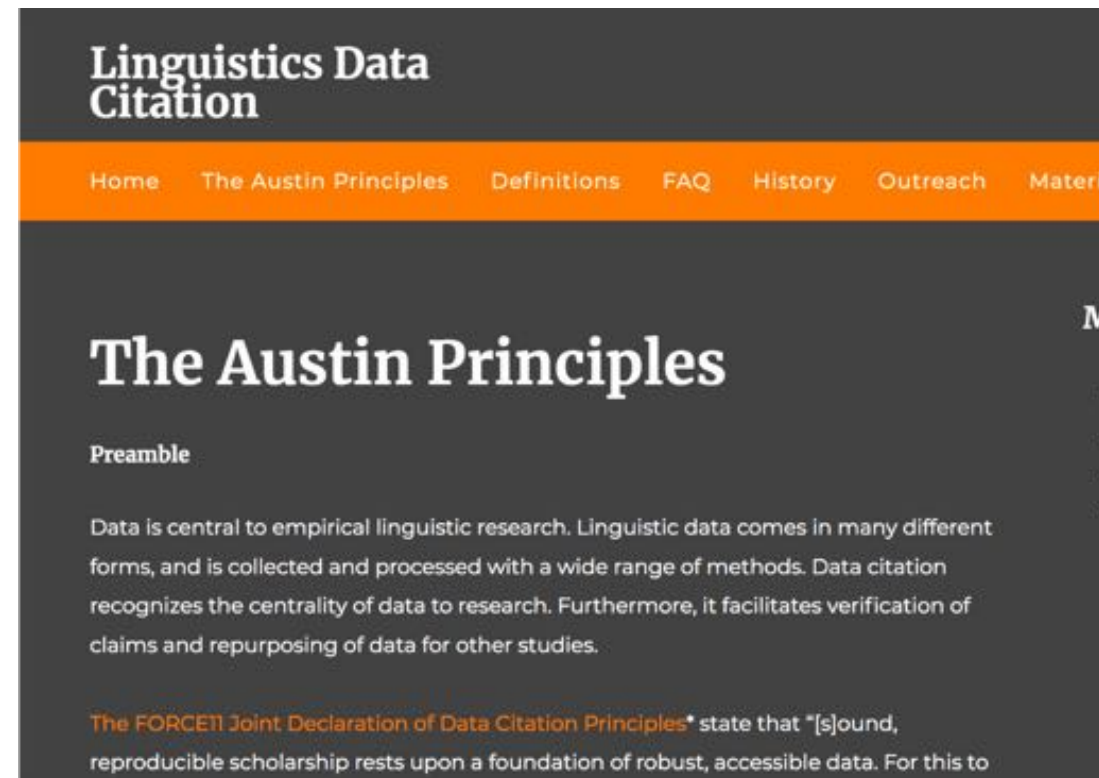
The Austin Principles of Data Citation in Linguistics

Output 1

Austin Principles of Data Citation in Linguistics

- Objectifs
 - Encourager et améliorer la visibilité et la récupérabilité des données.
 - Guidelines pour la mise en forme des citations de données.
 - Être applicable à toute sous-discipline ainsi que tout type de données.
- Modèle
 - FORCE11 Joint Declaration of Data Citation Principles (2014)

(Berez-Kroeker et al., 2018)



<https://site.uit.no/linguisticsdatacitation/>

Raisonnement

Pourquoi des principes spécifiques pour la linguistique?

- La linguistique, dans toute sa diversité, n'est pas nécessairement très différente des autres disciplines scientifiques, en ce qui concerne la gestion de données.
- L'idée principale a été de traduire les principes généraux, bien connus dans le monde RDM, dans un langage plus accessible au chercheur qui veut/doit apprendre les meilleures pratiques.

Austin Principles of Data Citation in Linguistics

1. Importance

En linguistique, les données ne sont **pas uniquement une source scientifique**: Elles peuvent également être une documentation de l'héritage culturel, de l'évolution sociétale et du potentiel humain.

A cause de cela, les données qui forment la base d'analyses linguistiques sont d'une **importance fondamentale** dans le domaine et devraient être traitées comme telles.

Les données linguistiques devraient être **citables et citées**, et ces citations devraient être traitées avec la même importance que les citations d'autres types de publication en linguistique.

(traduction non-officielle par l'auteure)

Austin Principles of Data Citation in Linguistics

2. Crédit et attribution

En linguistique, les citations devraient faciliter la tâche d'**identifier les personnes ayant contribué aux données**, et de quelle manière celles-ci ont contribué, lorsque approprié.

Une solution est de **lister les contributeurs et leur(s) rôle(s)** dans la citation. Une autre solution est d'utiliser la citation comme un lien aux métadonnées qui elles présentent les contributeurs et leur(s) rôle(s).

Austin Principles of Data Citation in Linguistics

3. Preuve

Les linguistes devraient **citer les données** qui forment la base d'arguments scientifiques.

Pour que les données puissent être citées, elles devraient être **stockées dans un endroit accessible**, préférablement dans une archive fiable.

Les auteurs devraient vérifier que **la collecte de données et la méthode de traitement sont transparentes**, soit à travers un lien aux métadonnées, soit à travers une explication dans le texte. Cela soulignerait le lien entre les données et les arguments qui reposent sur celles-ci.

Austin Principles of Data Citation in Linguistics

4. Identification unique

Lors de la sélection d'archive, qui va préserver et donner accès aux données, les linguistes devraient chercher **un service qui offre une identification** sous la forme d'un identifiant pérenne (PID).

Pour les données numériques, cela peut être p.ex. un DOI ou un Handle.

Austin Principles of Data Citation in Linguistics

5. Accès

Les données linguistiques devraient être **aussi ouvertes que possible, afin de faciliter la reproductibilité, et aussi fermées que nécessaire**, afin de respecter les contraintes éthiques, légales et communautaires.

Austin Principles of Data Citation in Linguistics

6. Pérennité

Les linguistes devraient assurer que l'archive où ils stockent leurs données ont une **politique écrite qui déclare la pérennité** des données, des métadonnées et des identifiants.

Austin Principles of Data Citation in Linguistics

7. Spécificité et vérifiabilité

La citation devrait faciliter la tâche de **recupérer une donnée spécifique ou un sous-ensemble** de données, sur lequel se base un argument scientifique.

Si l'usage des données nécessite une certaine granularité au niveau des citations, il faudrait avoir une **méthode systématique de récupération**.

Beaucoup d'ensembles de données ne sont pas statiques, p.ex. des ensembles où des données sont ajoutées régulièrement. Les citations devraient spécifier **quelle version** des données a été utilisée.

Austin Principles of Data Citation in Linguistics

8. Interopérabilité et flexibilité

Les linguistes travaillent sur **différents types de données**, et sur des questions de recherche très variées.

Les standards de citation développés pour la linguistique devraient refléter ce fait, sans aller contre les principes de citation généraux.

Nous encourageons les éditeurs de publications en linguistique de faciliter la tâche de citation de données et de développer des formats de citation, ainsi qu'une politique de citation non-ambiguë, sur la base de ce document.

Retour sur la V1 de la communauté LDIG

Linguistics Data Interest Group (LDIG)

Meeting - August 30-~~September 8~~ 15 2017

Pre-RDA Plenary 10

-- Document editing now closed --

How to participate

Welcome to this asynchronous meeting of the [Linguistics Data Interest Group](#) (LDIG).

There are three ways to participate:

1. Enter text or add a comment to this document. Make sure you include your name and list yourself as a meeting participant (with your email address) at the end of the document.
2. Email feedback to Lauren Gawne (l.gawne@latrobe.edu.au), which will be incorporated into the final version (anonymously if requested).
3. Add your feedback in [this Google Form](#), which asks the same questions as in this meeting document. Your feedback will be incorporated into the final version (anonymously if requested).

Aims of this meeting

1. Get feedback on the [Austin Principles of Data Citation](#)
2. Find specific examples with which to illustrate the Austin Principles
3. Get feedback on other issues and ideas to take to the RDA's 10th Plenary
4. Keep LDIG members up to date on what we are working on

Retour sur la V1 de la communauté LDIG

Feedback on the Principles (Aim 1)

What are your first impressions reading this document?

██████████: I am mostly quite pleased with the principles as laid out in the document, although I do think that they will be more easily applied to some sorts of research (e.g. archival data collection and cultural curation) than some others (e.g. syntactic, morphological, phonological analysis of linguistic data).

██████████ I am also quite pleased - I think the main thrust is that data underpinning any linguistic work should be available for replication, if nothing else, which I see as a positive thing. I think Stan's point is valid, but I would also say that analysis of linguistic data can be a dataset (of forms, sentences used in analysis, recordings of sounds, scripts used in R or Praat, Python code, or a host of other kinds of data). The main thing is that it is incumbent on the linguist to track and explain their particular method of reaching particular conclusions, which is just good science.

██████████: I concur. Building on Stan's point, I think that the principles work better when they are formulated in terms that suit the whole linguistics community than when they provide examples (e.g., the DELAMAN archives) that will only be familiar to people in specific subfields. I think it might make sense to reformulate #4 in that way, e.g., 'Data repositories, corpora and other resources for housing and providing access to linguistic data should provide the means for such identification in the form of a PID...'. [Incidentally, it seems that the TROLLing Dataverse is down; if that's permanent, it would be wise to remove it as an example]

██████████: I also concur. Stan's (good) point, perhaps, relates to questions some readers might have about what constitutes "data" for these principles. I realize this comes from FORCE11, but the preamble states "data should be considered legitimate, citable products of research" without explaining what "data" are, and whether the consideration is from the creator of the data's perspective or the users' or... But, I realize this comment drifts away from the principles itself and gets into some of the related muddy waters we've talked about in our meetings (e.g. how much of this is really just about citation practices vs larger principles about open-science-based approaches)... In sum, I really like the document and think it does a great job of hitting the main

Le travail approuvé par la communauté

Linguistics Data Citation

Home The Austin Principles Definitions FAQ History Outreach Materials Endorse

Endorse

Data citation will only become standard practice in linguistics if individual researchers and entire research communities commit to supporting data citation across the discipline.

By endorsing the Austin Principles you show other linguists that you support reproducible research. **Individuals** can join the list of linguists who endorse the principles by adding your name to the [Individual Endorsement Google Form](#). **Organizations** can join the list of endorsers by filling out the [Organization Endorsement Google Form](#).

If you would like to be more actively involved in shaping the future of linguistic data, we encourage you to join the [Linguistics Data Interest Group](#) as part of the Research Data Alliance.

META

- Log in
- Entries RSS
- Comments RSS
- WordPress.org

Organization Endorsements		Individual Endorsements			
	CLARIN - European Research Infrastructure for Language Resources and Technology	Mortensen	David R.	Carnegie Mellon University	April
		Han	Na-Rae	University of Pittsburgh	April
		Cox	Christopher	Carleton University	June
	IRIS: A Digital Repository of Instruments and Materials for Research Into Second Languages	Sammons	Olivia	University of Alberta	June
		Gasser	Emily	Swarthmore College	July 3
		Huddleston	Kate	Stellenbosch University	July 3
	Kaipuleohone University of Hawai'i Digital Language Archive	Rebecca	Whitney	Tulane University, Quapaw Tribe of Oklahoma	July 8
		Hellmuth	Sam	University of York	July 8
	Tromsø Repository of Language and Linguistics (TROLLing)	Fionn	Volkov	University of Queensland	July 3
		Tony	McEnery	Lancaster University	August
		Thom	van Hugte	Leiden University	August
	Archive of the Indigenous Languages of Latin America	Aidan	Solla	Monash University	August
		Taylor	Charlotte	University of Sussex	September
	Pacific and Regional Archive for Digital Sources in Endangered Cultures	Rathmann	Christian	Humboldt-Universität zu Berlin	September
		Bowen	Eliza	University of Western Australia	September

Recommendations for citation of research data in linguistics (working title)

Output 2

Raisonnement

Pourquoi des recommandations spécifiques pour la linguistique?

- A nouveau, la linguistique, dans toute sa diversité, n'est pas nécessairement très différente des autres disciplines scientifiques, en ce qui concerne la citation de données.
- Mais rappelons Austin Principe no. 8:
Les linguistes travaillent sur différents types de données, et sur des questions de recherche très variées.
Les standards de citation développés pour la linguistique devraient refléter ce fait, sans aller contre les principes de citation généraux.

Recommendations for citation of research data in linguistics (working title)

- Modèle de citation
 - Citation dans le texte
 - Référence bibliographique
 - Exemples commentés
 - Définitions élaborées
- Public visé
 - Editeurs de publications scientifiques (conseils aux auteurs)
 - Chercheurs qui vont citer des données et/ou qui vont archiver des données.
 - Archives de données scientifiques (conseils et exigences au niveau des métadonnées)

Recommendations for citation of research data in linguistics (working title)

The template for a **minimal reference to a dataset resource in the bibliography** section of a piece of academic writing is:

Author, Date, Title, Publisher, Locator.

The template for an **expanded bibliographic reference** to a dataset resource, including *conditional elements* (i.e. required in certain cases depending on resource characteristics) is:

Author, Date, Title, Publisher, Locator, Version, Date accessed, Tag.

In-text (or in-line) citations must point to a bibliographic reference at the end of the published work. The template for a **minimal in-text citation** is:

Author, Date

The template for an **expanded in-text citation** including additional potential information is:

Author, Date, Locator, Subset, Other Attribution (Roles)

3. Glossary

- **Author:** By "Author" is meant by default one or more entities (persons or organisations) responsible for developing the resource **and depositing it**. Specific **Roles** will vary with the details of the resource and the terminology of the resource provider, but might include "Project Leader", "Investigator", "Researcher", "Data Collector", "Project Contact," "Consultant," etc. By default, only the main responsible **Authors** are listed without mentioning their roles. Other entities might be specified or required by a resource provider, publisher guidelines, or subfield norms, and should be marked by their specific roles using parentheses, e.g. *John Smith (Data Collector)*. When roles may be needed in addition to Author, we have listed these as **Other Attribution** in our templates.

For more information on contributor roles, see the following:

- <https://www.casrai.org/credit.html>

-
- Definitions in the DataCite Metadata Schema Documentation for the Publication and Citation of Research Data, Version 4.3:
https://schema.datacite.org/meta/kernel-4.3/doc/DataCite-MetadataKernel_v4.3.pdf
 - OLAC Role vocabulary: <http://www.language-archives.org/REC/role.html>

Retour sur la V1 de la communauté

Linguistics Data Interest Group (LDIG)

Recommendations for citation of research data in linguistics

Draft 2

29 August 2019

-- document editing and commenting open --

This is the second draft of the recommendations, which reflects edits made after the RDA plenary during the [LDIG session](#) on Wednesday 3 April 2019. We invite everybody receiving the link to this document, to read, comment, and provide examples for various types of linguistic data, for various subfields in linguistics.

How to participate: Enter text or add a comment to this document. Make sure you include your name and list yourself as a contributor (preferably with your email address) at the end of the document.

Commentaires encouragés

- dans la communauté LDIG
- chez des éditeurs, directeurs, présidents d'organisations et comités, et d'autres personnes actives dans le mouvement Open Science

Retour sur la V1 de la

Prignitz, Gisèle. 2007 (collection date). Enquête Burkina Faso. Projet PFC.
<https://public.projet-pfc.net/>. (Accessed 2019-06-22).

The template for a **minimal reference to a dataset resource** in the **bibliography** section of a piece of academic writing is:

Author, Date, Title, Publisher, Locator.

The template for an **expanded bibliographic reference** to a dataset resource, including **conditional elements** (i.e. required in certain cases depending on resource characteristics) is:

Author, Date, Title, Publisher, Locator, Version, Date accessed, Tag.

In-text (or in-line) citations must point to a bibliographic reference at the end of the published work. The template for a **minimal in-text citation** is:

Author, Date

The template for an **expanded in-text citation** including additional potential information is:

Author, Date, Locator, Subset, Other Attribution (Roles)

Please note: Definitions of the elements contained in the bibliographic reference and the in-text citation can be found in the [Glossary](#). A longer version of the recommendations, explaining concepts, highlighting challenges and providing examples can be found in: Konzett, Philipp & Koenraad De Smedt. (in preparation). Guidance for citing research data. In Andrea L. Berez-Kroecker, Bradley McDonnell, Lauren Collister & Eve Koller (eds.), *Open Handbook of Linguistic Data Management*. MIT Press Open.

in item and the roles in all item are listed with several persons and their roles on the respective landing pages. With no reference recommended by the resource provider, we include all persons in the **Author** field, and also their roles. In these specific cases, Hauk and Krauss most closely fit the definition of **Author** (i.e., the person most responsible for developing the

9 PM Sep 25

sure that annotations on dates
d idea. Complex dates such as
22) or "1999[1945]" are often
ut "2007 (collection date)"
break the convention that
uld be integers without a very
ivation. If you do a lookup for
of a decade, for instance, this
ld not be retrieved.
that there is no "real" date of
in seems to be not a very
piece of information that
ept at all costs.

8:58 PM Sep 18

Delete: "a"

8:54 AM Sep 12

Add: "These recommendations are for the
use of researchers and scholars in the field
working with datasets..."

8:55 AM Sep 12

Sorry, that sentence was too long for my
tired brain

8:40 AM Sep 18

Delete space

9 AM Sep 17

experience, URI and URL, have
d interchangeably so often and
g, that any subtle differences
he two are lost on most
so from a usability/readability
ve, I'd recommend to stick with

8:32 AM Sep 17

This would be categorized as PID
following the terminology above, right?

Calendrier

Octobre 2019

Révision et finalisation des recommandations.

Janvier 2020

Publication avec DOI sous format de «RDA Supporting Output».

Participer à la discussion jusqu'au 14 octobre

<https://www.rd-alliance.org/groups/linguistics-data-ig>



<https://www.rd-alliance.org/recommendations-outputs>

The Open Handbook of Linguistic Data Management

Output 3

Open Handbooks In Linguistics

This blog catalogs developments in the Open Handbooks in Linguistics series, started in fall 2016 by a group of concerned linguists. The initiative is to create high-quality open-access linguistics handbooks which are accessible to all.

Monday, April 16, 2018

Series Contract Signed!!



Greetings everyone!

Long time no hear, I know, but the wheels of open access have been grinding on, albeit extremely slowly. And I am extremely pleased to announce that we now have a series contract in hand with MIT Press! Open Handbooks in Linguistics is officially a Thing!

The first volume in the series, edited by [Andrea Berez-Kroeker](#), [Bradley McDonnell](#) and [Eve Okura Koller](#), will be "The Open Handbook of Linguistic Data Management and Archiving." It looks like it is going to be "awesome".

We will be establishing a process for identifying the next volumes to be pursued for publication in the next month or two, and we'll make announcements accordingly, so please watch out for them!

Each volume in the series will require a financial subvention to support Open Access publication, approximately \$13,000 per volume. We will apply to various public and private sources of funding to support the series, but it will also be possible for individuals to make contributions in support of specific projects that they especially believe in.

EXTREMELY excited to finally be able to make this announcement,

your open access handbooks advocate,

Heidi Harley

Blog Archive

▼ 2018 (1)

▼ April (1)

[Series Contract Signed!!](#)

► 2016 (3)

Publication en cours

The Open Handbook of Linguistic Data Management

Open Handbooks in Linguistics series, MIT Press Open, 2020.

Andrea L. Berez-Kroeker, Bradley McDonnell, & Eve Koller (éds.)

- 13 chapitres sur le fondement conceptuel de la gestion de données de recherche en linguistique, ainsi que sur les meilleures pratiques.
- 50 exemples d'implémentation des meilleures pratiques.
- Plus de 70 contributeurs venant de quatre continents.

La prochaine étape de LDIG

Diffusion, éducation et sensibilisation

Les meilleures pratiques et les défis actuels

Les chercheurs établis

- Répondre aux exigences.
- Etablir de nouvelles routines.

Les jeunes chercheurs

- Acquérir une culture de transparence et de partage.

Les organismes qui développent les services de soutien

- Contribuer à l'éducation des (jeunes) chercheurs.
- Diffuser des informations sur les meilleures pratiques et les services développés.

Réflexion et invitation

- Le travail débuté en 2015 avec le projet NSF a montré que **la collaboration** entre différents acteurs est essentielle pour faire avancer le changement de culture.
- **Les meilleures pratiques** de la gestion des données en linguistique, qu'il s'agisse de la collecte de données, du développement d'archive ou du lien entre le texte et les données, **doivent être intégrées** dans le travail des différents acteurs dès maintenant.
- **Les exigences sont déjà là** (institutions, bailleurs de fonds, revues), reste le travail qui va faire de la recherche en linguistique une recherche plus transparente et de ce fait, une recherche de meilleure qualité.
- **Le groupe LDIG invite** désormais toute personne présente dans cette salle à joindre la communauté RDA afin de contribuer au projet avec vos expériences et compétences.



La citation de données de recherche dans des publications en linguistique

– l’histoire d’une initiative prise par
la communauté scientifique

helene.n.andreassen@uit.no

Journée évaluation des corpus : où en est-on, où allons-nous ?

Séminaire organisé par CORLI, 3 octobre 2019, Paris

Cette présentation illustre un travail fait en collaboration avec Andrea Berez-Kroeker (U Hawai’i at Manoa), Lauren Gawne (La Trobe U), Lauren Collister (U Pittsburgh), Christopher Cox (Carleton U), Philipp Konzett (UiT), Koenraad De Smedt (U Bergen) & Bradley McDonnell (U Hawai’i at Manoa)



Bibliographie

- Berez-Kroeker, A. L., Andreassen, H. N., Gawne, L. , Holton, G., Kung, S. S., Pulsifer, P., Collister, L. B., The Data Citation and Attribution in Linguistics Group, & The Linguistics Data Interest Group. (2018). *The Austin Principles of Data Citation in Linguistics*. V1.0. <http://site.uit.no/linguisticsdatacitation/austinprinciples/>. Accessed 01.10.2019.
- Berez-Kroeker, A. L., Gawne, L., Kelly, B. F., & Heston, T. (2017). *A survey of current reproducibility practices in linguistics journals, 2003-2012*. <https://sites.google.com/a/hawaii.edu/data-citation/survey>.
- Berez-Kroeker, A. L., Gawne, L., Kung, S. S., Kelly, B. F., Heston, T., Holton, G., . . . Woodbury, A. C. (2018). Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics*, 56(1), 1-18. <https://doi.org/10.1515/ling-2017-0032>
- Data Citation Synthesis Group, & Martone, M. (ed.). (2014). *Joint Declaration of Data Citation Principles*. San Diego CA: FORCE11. <https://doi.org/10.25490/a97f-egyk>
- Gawne, L., Kelly, B. F., Berez-Kroeker, A. L., & Heston, T. (2017). Putting practice into words: The state of data and methods transparency in grammatical descriptions. *Language Documentation & Conservation*, 11, 157-189. <http://hdl.handle.net/10125/24731>
- Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Lucraft, M., & Astell, M. (2018). *Practical challenges for researchers in data sharing*. <https://doi.org/10.6084/m9.figshare.5975011>
- Thomason, S. G. (1994). The Editor's Department. *Language*, 70(2), 409-413. <http://www.jstor.org/stable/415877>
- Wouters, P., & Haak, W. (2017). *Open data: The researcher perspective*. <https://www.elsevier.com>