

# Multilinguisme et multiplicité des langues: État des lieux et défis pour le TAL

Laurent Besacier<sup>1</sup> et François Yvon<sup>2</sup>

à partir d'un rapport rédigé avec M.-R. Amini, L. Barrault, D. Bernhard,  
O. Kraif, E. Morin et K. Smaïli

(<sup>1</sup>) LIG, CNRS et Univ. Grenoble Alpes et (<sup>2</sup>) LIMSI, CNRS et Univ. Paris-Saclay

Grenoble, 24 octobre 2019

Etat des lieux

Propositions

# Contexte

- ▶ Des ressources langagières pour un nombre croissant de langues
- ▶ Nourrissent des approches véritablement multilingues
  - ▶ en traduction automatique [Koehn, 2005],
  - ▶ en analyse syntaxique [Nivre et al., 2016],
  - ▶ en reconnaissance automatique de la parole [Schultz et al., 2013, Schultz and Schlippe, 2014],
  - ▶ en désambiguisation lexicale [Navigli and Ponzetto, 2010, Sérasset, 2014],
  - ▶ en dialectologie computationnelle [Christodoulopoulos and Steedman, 2015],
  - ▶ etc.

# Périmètre de l'axe

- ▶ Traduction automatique de l'écrit et de l'oral
- ▶ Alignement, transfert et similarité entre langues
- ▶ Traitement automatique des dialectes, des langues régionales et des langues peu dotées
- ▶ TAL et apprentissage des langues
- ▶ TAL et dialectologie
- ▶ Modélisation des phénomènes d'alternance codique (*code switching*)

# Traduction automatique de l'écrit et de l'oral

- ▶ Avancées tangibles ces dernières années [Bahdanau et al., 2015, Vaswani et al., 2017] mais ...
- ▶ Long chemin à accomplir pour atteindre une traduction automatique **de haute-qualité, transparente et fiable**
  - ▶ Pour tous les couples de langues, domaines et contextes de traduction
  - ▶ Selon des scénarios d'usage multiples : traduction automatique, traduction assistée, traduction de contenus bruités, traduction de l'oral, de langue des signes, etc

# Traduction automatique de l'écrit et de l'oral

- ▶ L'addiction aux données :
  - ▶ corpus parallèles trop petits, trop spécialisés, hétérogènes mais insuffisamment divers
  - ▶ disponibles pour un petit nombre de styles, registres, domaines, et paires de langues
  - ▶ foisonnement de travaux autour de l'adaptation au domaine [Chu et al., 2017]; la traduction multilingue [Johnson et al., 2016b]; l'apprentissage sans données parallèles [He et al., 2016, Artetxe et al., 2018, Lample et al., 2018]
- ▶ Un recours ? Utiliser des connaissances / contraintes expertes ou contextuelles
  - ▶ syntaxe en langue source [Aharoni and Goldberg, 2017, Bastings et al., 2017, Li et al., 2017, Tran and Bisk, 2018]
  - ▶ contextes longs (paragraphe, document) nécessaire pour co-références [Bawden et al., 2018], ellipse, etc
  - ▶ cohérence stylistique du discours [Tiedemann and Scherrer, 2017, Jean et al., 2017],

Encore peu de travaux sur l'intégration de connaissances sémantiques ou structure du discours

# Traduction automatique de l'écrit et de l'oral

- ▶ Gap sémantique
  - ▶ Traduction de mots sources sémantiquement ambigus implique (?) d'identifier le sens correct en contexte
  - ▶ Mais les modèles actuels emploient des techniques robustes et n'analysent pas le sens de la phrase source
  - ▶ Méthodes issues de la sémantique computationnelle encore peu utilisées dans les systèmes de TA
  - ▶ Perspective :
    - ▶ meilleure intégration des approches de désambiguïsation lexicale en TA [Gonzales et al., 2017]
    - ▶ traduction de représentations sémantiques en source [Song et al., 2019]
    - ▶ utilisation d'images comme forme d'ancrage sémantique (*grounding*) de la TA [Barrault et al., 2018]

# Traduction automatique de l'écrit et de l'oral

- ▶ Traduction de la parole
  - ▶ Quelle(s) représentation(s) intermédiaire(s) (orthographique et—ou /fonetik/) pour l'énoncé source ?
  - ▶ Traduction directe « bout en bout » [Berard et al., 2016, Weiss et al., 2017, Bérard et al., 2018]
  - ▶ Intérêt pour la traduction de dialectes oraux et pour la traduction de parole en général (reduire temps de décodage et empreinte mémoire des modèles)
  - ▶ Vers l'**interprétation** simultanée ? [Ma et al., 2018]



# Alignement, Transfert, Similarités

- ▶ Alignement de corpus parallèles
  - ▶ Tâche intégralement non-supervisée, mal définie, mais...
  - ▶ Applications variées : extraction de lexiques [Morin and Daille, 2012], recherche d'information interlingue [Nie, 2010], documentation de langues peu dotées [Godard et al., 2016, Anastasopoulos and Chiang, 2017], apprentissage des langues, estimation de la qualité [Specia et al., 2017], etc.
  - ▶ Exploitation de corpus comparables [Zweigenbaum and Benoît, 2006] ou multimodaux [Robert-Ribes et al., 1997, Toselli et al., 2011]

# Alignement, Transfert, Similarités

- ▶ Transfert
  - ▶ Projeter rapidement des modèles de TAL d'une langue à une autre,
  - ▶ Mécanisme peu coûteux pour exploiter des ressources annotées disponibles dans d'autres langues
  - ▶ Appliqué à l'analyse morpho-syntaxique [Yarowsky et al., 2001, Wisniewski et al., 2014, Zennaki et al., 2016], syntaxique [Hwa et al., 2005, Tiedemann, 2014, Lacroix et al., 2016, Aufrant et al., 2016] ou sémantique [Padó and Lapata, 2009, Jabaian et al., 2013]
  - ▶ Exploitation de corpus parallèles et/ou mutualisation de parties du modèle (neuronal) [Ha et al., 2016, Johnson et al., 2016a, Gu et al., 2018].
  - ▶ Un enjeu : définir des représentations communes à toutes les langues (*interlingua*?)

# Alignement, Transfert, Similarités

- ▶ Construction de plongements interlingues
  - ▶ Représentations **indépendantes des langues** d'unités linguistiques (vocables, phrases)
  - ▶ Similarités cross-lingues  $\approx$  traductions
  - ▶ Applications : détection de plagiat translingue [Schwab et al., 2017], recherche d'information multilingue [Balikas et al., 2018], implication textuelle translingue [Conneau et al., ], TA (!) etc.

# Langues peu dotées, dialectes, langues régionales

- ▶ Variabilité, manque de données, collecte coûteuse
- ▶ Besoin de méthodes robustes et frugales en données, par ex : transfert (de modèles ou d'annotations)
- ▶ Depuis quelle(s) langue(s) source(s) et vers quelle(s) langue(s) cibles transférer ?
- ▶ Travaux actuels exploitant des données typologiques (par ex. WALS<sup>1</sup>)
- ▶ Approches non-supervisées correspondant à la tâche (*zero shot MT*)
  - ▶ En parole, recherches initiées via le défi *Zero Resource Speech Challenge* [Dunbar et al., 2017]<sup>2</sup>
  - ▶ Lien avec l'acquisition du langage chez l'enfant

---

1. <https://wals.info>

2. <http://zerospeech.com>

# Animation de la communauté

- ▶ État de l'art partagé
- ▶ Journées scientifiques thématiques (sur le modèle des journées ATALA)
- ▶ Groupes de lecture (potentiellement à distance) - bourse de voyage contre billet post-conférence
- ▶ Défis et tâches partagées (traduction de langues proches, de langues régionales)
- ▶ École d'hiver en préparation (L. Besacier et M. Gallé) pour janvier 2021 dans le Vercors
  - ▶ 4 invités déjà confirmés : Tim Baldwin, Kyunghyun Cho, Claire Gardent, Sanjeev Khudanpur

Implication des acteurs non-académiques ?

# Susciter des actions communes

vers d'autres points de fonctionnement des systèmes de traduction automatique

Pour poser de nouvelles questions en terme de défis, d'architecture et d'évaluation

- ▶ Langues proches ( $BLEU > 60$ )
  - ▶ Traduction Français-Espagnol-Italien ?
  - ▶ Analyse d'erreurs
- ▶ Domaine littéraire ( $BLEU < 15$ )
  - ▶ Traduction de métaphores et d'autres procédés littéraires ?
  - ▶ Traduction d'énonciations humoristiques ?
- ▶ Parole expressive
  - ▶ Expression des émotions
  - ▶ « Traduction » de la prosodie

# Susciter des actions communes (2)

Développer / exploiter des ressources bilingues français / langues régionales

- ▶ Breton, Picard, Alsacien, Corse, Catalan, Basque, Créoles, etc.
  - ▶ Ressources alignées avec le français
    - ▶ Pour la traduction automatique
    - ▶ Pour la projection translingue d'annotations / modèles
- Enjeu : outillage (par transfert) de langues peu dotées
- ▶ Diffusion vers la communauté
    - ▶ Organisation d'un challenge / Hackathon?

# Susciter des actions communes (2)

Développer / exploiter des ressources bilingues français / langues régionales

- ▶ Breton, Picard, Alsacien, Corse, Catalan, Basque, Créoles, etc.
  - ▶ Ressources alignées avec le français
    - ▶ Pour la traduction automatique
    - ▶ Pour la projection translingue d'annotations / modèles
- Enjeu : outillage (par transfert) de langues peu dotées
- ▶ Diffusion vers la communauté
    - ▶ Organisation d'un challenge / Hackathon?

Idem pour les langues de migration



# Interactions au sein du GDR

- ▶ **Apprentissage et modèles** : Apprendre avec peu de données - et des modèles / connaissances, adaptation au domaine, apprentissage structuré, apprentissage des structures, apprentissage multitâche, apprentissage zero-shot, représentations multilingues et transfert.
- ▶ **Multimodalité, etc** Modéliser des contextes longs, ancrage sémantique, utilisation des représentations sémantiques profondes ? Analyser / reconnaître / engendrer de l'expressivité, traduire entre modalités (y-c langue des signes)
- ▶ **Accès à l'information** Modèles de transfert X-lingue pour l'accès à l'information ?
- ▶ **Ressources** Quelles ressources pour la TA depuis / vers le français ?

# Encourager les interactions avec d'autres communautés

- ▶ Apprentissage automatique
- ▶ Traitement de la parole
- ▶ Traductologie et linguistique de corpus
- ▶ Terminologie et Ontologie
- ▶ *Computational Language Documentation* (planche suivante)
- ▶ Apprentissage des langues assisté par ordinateur
- ▶ Interaction Humain-Machine (outils de TAO / PE, interfaces de lecture / écriture bilingue, gestion de l'interaction multilingue)
- ▶ Vision artificielle?
- ▶ Neurosciences computationnelles (étude du bi/multi-linguisme)?

# Produire et diffuser de l'expertise

- ▶ Cours, Vulgarisation, Interventions publiques
- ▶ Posts de blog (notes de lectures, commentaires d'articles)
- ▶ Fil twitter ? autres ?

# Lien avec GdR LIFT (INSHS CNRS)

- ▶ LIFT : favoriser interactions entre linguistique informatique, linguistique formelle et linguistique de terrain
  - ▶ Extraction de généralisations linguistiques par des méthodes informatiques
  - ▶ Linguistique et évaluation des systèmes de traitement automatique des langues
  - ▶ Outils de collecte et d'analyse pour les linguistes
  - ▶ Données et défis partagés pour une science ouverte
  - ▶ Linguistique informatique pour les langues peu dotées ou non documentées
- ▶ Journées du GDR LIFT à Orléans les 28-29 novembre 2019

# References I



Aharoni, R. and Goldberg, Y. (2017).  
Towards string-to-tree neural machine translation.  
*In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2 : Short Papers)*, pages 132–140, Vancouver, Canada.  
Association for Computational Linguistics.



Anastasopoulos, A. and Chiang, D. (2017).  
A case study on using speech-to-translation alignments for language  
documentation.  
pages 170–178. Association for Computational Linguistics.



Artetxe, M., Labaka, G., Agirre, E., and Cho, K. (2018).  
Unsupervised neural machine translation.  
*In Proceedings of the Sixth International Conference on Learning  
Representations*.



Aufrant, L., Wisniewski, G., and Yvon, F. (2016).  
Zero-resource dependency parsing : Boosting delexicalized cross-lingual transfer  
with linguistic knowledge.  
*In Proceedings of COLING 2016, the 26th International Conference on  
Computational Linguistics : Technical Papers*, pages 119–130, Osaka, Japan.  
The COLING 2016 Organizing Committee.

# References II



Bahdanau, D., Cho, K., and Bengio, Y. (2015).  
Neural Machine Translation by Jointly Learning to Align and Translate.  
In *ICLR 2015*, pages 3104–3112, San Diego, California, USA.



Balikas, G., Laclau, C., Redko, I., and Amini, M. (2018).  
Cross-lingual document retrieval using regularized wasserstein distance.  
In *Advances in Information Retrieval - 40th European Conference on IR Research, ECIR 2018, Grenoble, France, March 26-29, 2018, Proceedings*, pages 398–410.



Barrault, L., Bougares, F., Specia, L., Lala, C., Elliott, D., and Frank, S. (2018).  
Findings of the third shared task on multimodal machine translation.  
In *Proceedings of the Third Conference on Machine Translation, Volume 2 : Shared Task Papers*, pages 308–327, Belgium, Brussels. Association for Computational Linguistics.



Bastings, J., Titov, I., Aziz, W., Marcheggiani, D., and Simaan, K. (2017).  
Graph convolutional encoders for syntax-aware neural machine translation.  
In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1957–1967, Copenhagen, Denmark. Association for Computational Linguistics.

# References III



Bawden, R., Sennrich, R., Birch, A., and Haddow, B. (2018).  
Evaluating discourse phenomena in neural machine translation.  
In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 1 (Long Papers)*, pages 1304–1313. Association for Computational Linguistics.



Bérard, A., Besacier, L., Kocabiyikoglu, A. C., and Pietquin, O. (2018).  
End-to-End Automatic Speech Translation of Audiobooks.  
In *ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, Alberta, Canada.



Berard, A., Pietquin, O., Servan, C., and Besacier, L. (2016).  
Listen and translate : A proof of concept for end-to-end speech-to-text translation.  
In *NIPS Workshop on End-to-end Learning for Speech and Audio Processing*, Barcelona, Spain.



Christodoulopoulos, C. and Steedman, M. (2015).  
A massively parallel corpus : the bible in 100 languages.  
*Language Resources and Evaluation*, 49(2) :375–395.

# References IV



Chu, C., Dabre, R., and Kurohashi, S. (2017).

An empirical comparison of domain adaptation methods for neural machine translation.

In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2 : Short Papers)*, pages 385–391. Association for Computational Linguistics.



Conneau, A., Rinott, R., Lample, G., Williams, A., Bowman, S. R., Schwenk, H., and Stoyanov, V.



Dunbar, E., Cao, X. N., Benjumea, J., Karadayi, J., Bernard, M., Besacier, L., Anguera, X., and Dupoux, E. (2017).

**THE ZERO RESOURCE SPEECH CHALLENGE 2017.**

In *IEEE Automatic Speech Recognition and Understanding (ASRU)*, Okinawa, Japan.








Godard, P., Adda, G., Adda-Decker, M., Allauzen, A., Besacier, L., Bonneau-Maynard, H., Kouarata, G.-N., Löser, K., Riolland, A., and Yvon, F. (2016).

Preliminary Experiments on Unsupervised Word Discovery in Mboshi.

In *Interspeech 2016*, San Francisco, California, USA.



# References V

-  Gonzales, A. R., Mascarell, L., and Sennrich, R. (2017). Improving word sense disambiguation in neural machine translation with sense embeddings. In *WMT*, pages 11–19. Association for Computational Linguistics.
-  Gu, J., Hassan, H., Devlin, J., and Li, V. O. K. (2018). Universal neural machine translation for extremely low resource languages. *CoRR*, abs/1802.05368.
-  Ha, T.-L., Niehues, J., and Waibel, A. (2016). Toward multilingual neural machine translation with universal encoder and decoder. *arXiv preprint arXiv :1611.04798*.
-  He, D., Xia, Y., Qin, T., Wang, L., Yu, N., Liu, T., and Ma, W.-Y. (2016). Dual learning for machine translation. In *Advances in Neural Information Processing Systems*, pages 820–828.
-  Hwa, R., Resnik, P., Weinberg, A., Cabezas, C., and Kolak, O. (2005). Bootstrapping parsers via syntactic projection across parallel texts. *Nat. Lang. Eng.*, 11(3) :311–325.

# References VI



Jabaian, B., Besacier, L., and Lefèvre, F. (2013).  
Comparison and combination of lightly supervised approaches for language portability of a spoken language understanding system.  
*IEEE Trans. Audio, Speech & Language Processing*, 21(3) :636–648.



Jean, S., Lauly, S., Firat, O., and Cho, K. (2017).  
Does Neural Machine Translation Benefit from Larger Context ?  
*ArXiv e-prints*.



Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., et al. (2016a).  
Google’s multilingual neural machine translation system : enabling zero-shot translation.  
*arXiv preprint arXiv :1611.04558*.



Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F. B., Wattenberg, M., Corrado, G., Hughes, M., and Dean, J. (2016b).  
Google’s multilingual neural machine translation system : Enabling zero-shot translation.  
*CoRR*, abs/1611.04558.



Koehn, P. (2005).  
Europarl : A Parallel Corpus for Statistical Machine Translation.  
*In Conference Proceedings : the tenth Machine Translation Summit*, pages 79–86, Phuket, Thailand. AAMT, AAMT.

# References VII



Lacroix, O., Aufrant, L., Wisniewski, G., and Yvon, F. (2016).  
Frustratingly easy cross-lingual transfer for transition-based dependency parsing.  
In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, NAACL, pages 1058–1063, San Diego, California.



Lample, G., Conneau, A., Denoyer, L., and Ranzato, M. (2018).  
Unsupervised machine translation using monolingual corpora only.  
In *Proceedings of the Sixth International Conference on Learning Representations*.



Li, J., Xiong, D., Tu, Z., Zhu, M., Zhang, M., and Zhou, G. (2017).  
Modeling source syntax for neural machine translation.  
In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, pages 688–697, Vancouver, Canada.  
Association for Computational Linguistics.



Ma, M., Huang, L., Xiong, H., Liu, K., Zhang, C., He, Z., Liu, H., Li, X., and Wang, H. (2018).  
STACL : simultaneous translation with integrated anticipation and controllable latency.  
*CoRR*, abs/1810.08398.

# References VIII



Morin, E. and Daille, B. (2012).

Revising the compositional method for terminology acquisition from comparable corpora.

In *Proceedings of COLING 2012*, pages 1797–1810, Mumbai, India. The COLING 2012 Organizing Committee.



Navigli, R. and Ponzetto, S. P. (2010).

Babelnet : Building a very large multilingual semantic network.

In Hajic, J., Carberry, S., and Clark, S., editors, *ACL*, pages 216–225. The Association for Computer Linguistics.



Nie, J.-Y. (2010).

Cross-language information retrieval.

*Synthesis Lectures on Human Language Technologies*, 3(1) :1–125.



Nivre, J., de Marneffe, M., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., McDonald, R. T., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., and Zeman, D. (2016).

Universal dependencies v1 : A multilingual treebank collection.

In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*.



Padó, S. and Lapata, M. (2009).

Cross-lingual annotation projection of semantic roles.

*J. Artif. Int. Res.*, 36(1) :307–340.

# References IX



Robert-Ribes, J., Mukhtar, R. G., and Crc, A. C. S. (1997).  
Automatic generation of hyperlinks between audio and transcript.  
In *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech '97)*, Rhodes, Greece.



Schultz, T. and Schlippe, T. (2014).  
Globalphone : Pronunciation dictionaries in 20 languages.  
In *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014.*, pages 337–341.



Schultz, T., Vu, N. T., and Schlippe, T. (2013).  
Globalphone : A multilingual text & speech database in 20 languages.  
In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26-31, 2013*, pages 8126–8130.



Schwab, D., Besacier, L., Ferrero, J., and Agnès, F. (2017).  
Using word embedding for cross-language plagiarism detection.  
In *EACL (2)*, pages 415–421. Association for Computational Linguistics.



Sérasset, G. (2014).  
DBnary : Wiktionary as a Lemon-Based Multilingual Lexical Resource in RDF.  
*Semantic Web – Interoperability, Usability, Applicability*, pages –.  
To appear.

# References X



Song, L., Gildea, D., Zhang, Y., Wang, Z., and Su, J. (2019).  
Semantic neural machine translation using AMR.  
*Transactions of the Association for Computational Linguistics*, 7 :19–31.



Specia, L., Scarton, C., and Paetzold, G. (2017).  
*Quality estimation for Machine Translation*.  
Synthesis Lectures on Human Language Technologies. Morgan & Claypool  
Publishers.



Tiedemann, J. (2014).  
Rediscovering annotation projection for cross-lingual parser induction.  
In *Proceedings of COLING 2014, the 25th International Conference on  
Computational Linguistics : Technical Papers*, pages 1854–1864, Dublin, Ireland.  
Dublin City University and Association for Computational Linguistics.



Tiedemann, J. and Scherrer, Y. (2017).  
Neural machine translation with extended context.  
In *Proceedings of the Third Workshop on Discourse in Machine Translation*,  
pages 82–92. Association for Computational Linguistics.



Toselli, A. H., Romero, V., and Vidal, E. (2011).  
Alignment between text images and their transcripts for handwritten documents.  
In Sporleder, C., van den Bosch, A., and Zervanou, K., editors, *Language  
Technology for Cultural Heritage*, pages 23–37, Berlin, Heidelberg. Springer  
Berlin Heidelberg.

# References XI



Tran, K. and Bisk, Y. (2018).  
Inducing grammars with and for neural machine translation.



Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017).  
Attention is all you need.  
*In Advances in Neural Information Processing Systems*, pages 6000–6010.



Weiss, R. J., Chorowski, J., Jaitly, N., Wu, Y., and Chen, Z. (2017).  
Sequence-to-sequence models can directly transcribe foreign speech.  
*arXiv preprint arXiv :1703.08581*.



Wisniewski, G., Pécheux, N., Gahbiche-Braham, S., and Yvon, F. (2014).  
Cross-lingual part-of-speech tagging through ambiguous learning.  
*In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1779–1785, Doha, Qatar.



Yarowsky, D., Ngai, G., and Wicentowski, R. (2001).  
Inducing multilingual text analysis tools via robust projection across aligned corpora.  
*In Proceedings of the First International Conference on Human Language Technology Research, HLT '01*, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

# References XII



Zennaki, O., Semmar, N., and Besacier, L. (2016).

Inducing multilingual text analysis tools using bidirectional recurrent neural networks.

In *COLING*, pages 450–460. ACL.



Zweigenbaum, P. and Benoît, H. (2006).

Faire se rencontrer les parallèles : Regards croisés sur l'acquisition lexicale monolingue et multilingue.