

L'acquisition de données orales de masse en contexte naturel : Le système LENA[®]

Mélanie CANAULT – Equipe DENDY – DDL



Un travail collaboratif



Marie-Thérèse LeNormand
INSERM & Laboratoire Psychopathologie et Processus de
Santé, Université Paris Descartes, Sorbonne, Paris Cité, France
LPP, Institut de Psychologie, 71 avenue Edouard Vaillant,
92100 Boulogne Billancourt, France



Hung Thai Van
Service d'Audiologie & d'Explorations Orofaciales, Hospices
Civils de Lyon
Centre Recherche en Neurosciences de Lyon, INSERM,
U1028, CNRS UMR 5292, Lyon, France



Plan

1. La problématique de l'acquisition et du traitement des données orales chez le jeune enfant
 - a. Les données en contexte naturel (Naturalistic Data)
 - b. Les gros échantillons (Big Data)

2. Le système LENA[®]
 - a. Qu'est-ce que c'est et comment ça marche ?
 - b. La congruence Homme – Machine
 - Un système conçu pour l'anglais
 - Des validations dans d'autres langues
 - Focus sur le français

3. Les nouvelles pistes



La problématique de l'acquisition et du traitement des données orales chez le jeune enfant





Les données en contexte naturel



Moyen d'investigation privilégié pour la collecte de données orales (audio et vidéo) chez le jeune enfant.



Les données en contexte naturel

Dispositif
d'enregistrement
(fixe - mobile)



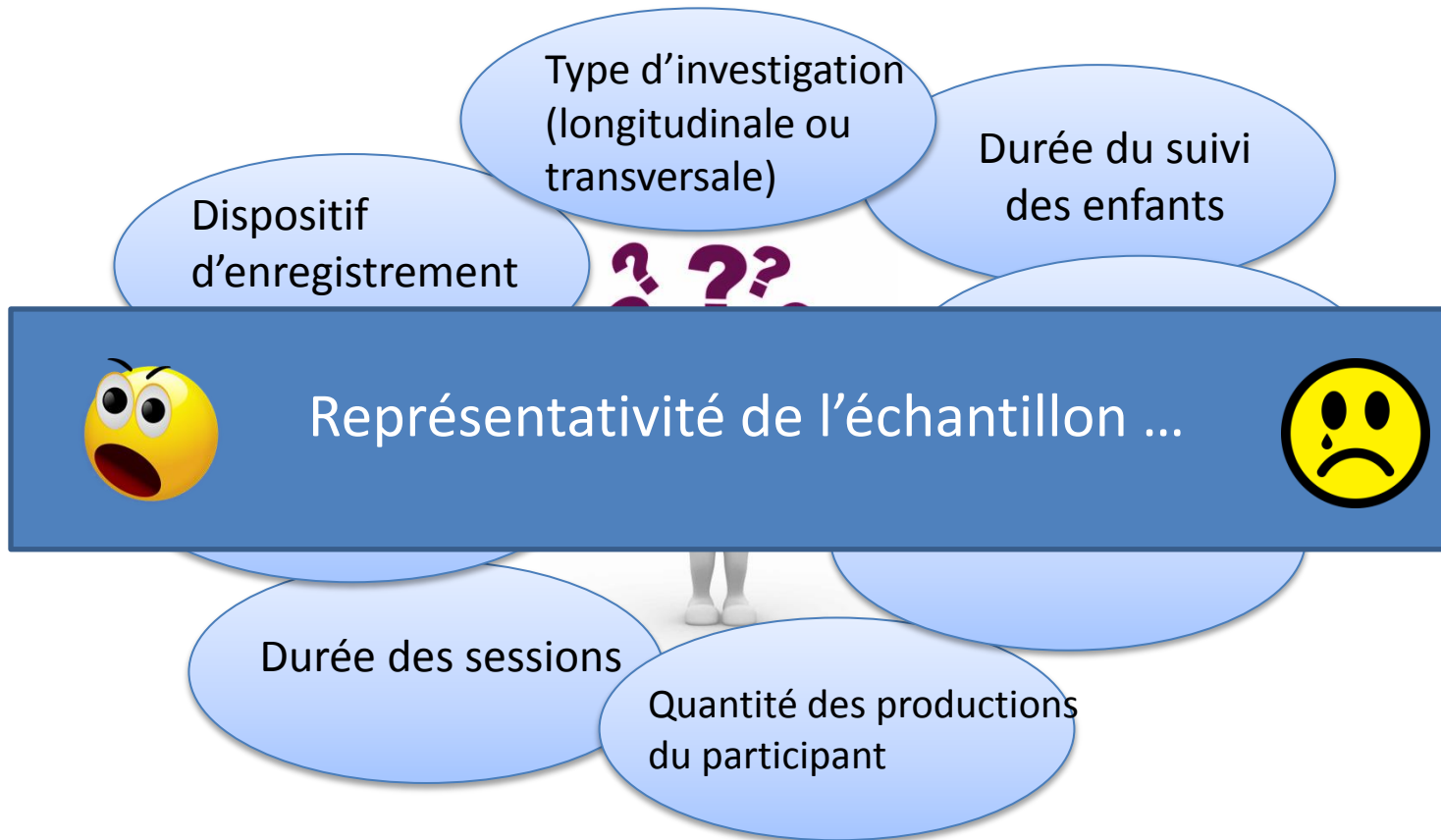


Les données en contexte naturel





Les données en contexte naturel





Les gros échantillons (Big Data)

L'environnement naturel ne suffit pas : besoin de gros échantillons (Big data)

- Le travail collaboratif des bases de données

Ex : CHILDES (MacWhinney, 2000) <http://childes.psy.cmu.edu/>

PhonBank : <http://phonbank.talkbank.org/> (Rose & MacWhinney, 2014 ; Rose, 2012)



Les gros échantillons (Big Data)

L'environnement naturel ne suffit pas : besoin de gros échantillons (Big data)

- Le travail collaboratif des bases de données

Ex : CHILDES (MacWhinney, 2000) <http://childes.psy.cmu.edu/>

PhonBank : <http://phonbank.talkbank.org/> (Rose & MacWhinney, 2014 ; Rose, 2012)



Mais parfois on n'y trouve pas ce dont on a besoin !



Les gros échantillons (Big Data)

Collecte de gros échantillons

Wells (1979) – le développement linguistique du jeune enfant (observation des interactions) : 128 enfants suivis plus de deux ans tous les 3 mois. Système d'Enregistrement conçu pour recueillir des séquences de 90 secondes toutes les 20 minutes de 9h à 18h.

Korman (1984) – les caractéristiques des productions orales maternelles adressées à l'enfant. Enregistrement de 24 heures tous les 2 mois : 6 familles d'enfants âgés de 6 à 16 semaines.

Hart & Risley (1995) – input et réussite scolaire. 42 familles suivies depuis les 7 mois jusqu'aux 36 mois de leur enfant enregistrées 1heure par mois. Soit un total de 1318 heures intégralement retranscrites.



Les gros échantillons (Big Data)

Collecte de gros échantillons

Wells (1979) – le développement linguistique du jeune enfant (observation des interactions) : 128 enfants suivis pendant 90 semaines.

Korfmeyer (1979) – Enregistrement de l'interaction parent-enfant.



Etudes rares et difficiles à répliquer
Faisabilité ...



Hart & Risley (1995) – input et réussite scolaire. 42 familles suivies depuis les 7 mois jusqu'aux 36 mois de leur enfant enregistrées 1heure par mois. Soit un total de 1318 heures intégralement retranscrites.



Le système LENA[®] : Language **EN**vironment **A**nalysis system



info@lenafoundation.org



Le système LENA[®]

Fonctionnalités de LENA

1. Enregistreur



(Digital Language Processor DLP)

Facile d'utilisation – Peu encombrant (petit, sans fil)

Autonomie de 16h

Bonne qualité acoustique



Le système LENA[®]

Fonctionnalités de LENA

2. Transfert des données



Software

The screenshot displays the LENA Pro software interface. At the top, there are navigation icons for Client Manager, LENA Reports, Developmental Snapshot, Automatic Vocalization Assessment, Digital Language Processor, and Settings. The main window is divided into several sections:

- Client Manager:** A table listing clients with columns for Client/Group, ID, and Notes. The data is as follows:

Client/Group	ID	Notes
Study 1	G01	collect audio
Study 2	G02	tv intervention
Study 3	G03	acoustic signal
Study 4	G01	ava results
- Child:** A table listing children with columns for Last Name, First Name, Birth Date, G, Start, ID, Child Key, DLP, and Notes. The data is as follows:

Last Name	First Name	Birth Date	G	Start	ID	Child Key	DLP	Notes
Ainsworth	James	2004-05-22	M	2008-10-13	CO07	48P3ACE70000295R	555	audio
Davis	Stacey	2003-11-23	F	2008-10-13	CO06	48P38CF80000F97R	655	signal level
Mullens	Denise	2005-12-21	F	2008-10-23	IP6_055	4900A1740000Q35R	3925	delayed
Sarens	Carter	2004-10-15	M	2008-10-09	IP1_900	48EE5E840000C00T	3016	average
Simms	Nacale	2004-12-14	F	2008-10-24	IP1_004	4901ESF90000F55R	7422	advanced
Wilson	Cooper	2003-07-23	M	2008-10-23	IP1_845	4900A1530000Q35R		tv
- Client / Group Information:** Fields for Name, ID, and Notes.
- Child Information:** Fields for First Name (Nacale), Last Name (Simms), Birth Date (12/14/2004), Gender (M), Child Key (4901ESF90000F55R), ID (IP1_004), DLP (7422), Client/Group, Start Date (2008-10-24), End Date, and Notes (advanced).



Le système LENA[®]

Fonctionnalités de LENA

3. Traitement des données

- 3h à 16h
- Pré-segmentation préalable

Localisation et codage des différentes portions de signal

Vocalisations de l'enfant : CHN/CHF

Productions de l'homme adulte : MAN/MAF

Productions de la femme adulte : FAN/FAF

Chevauchement : OLN/OLF

TV/Média : TVN/TVF

Bruit : NON/NOF

Autre enfant : CXN/CXF

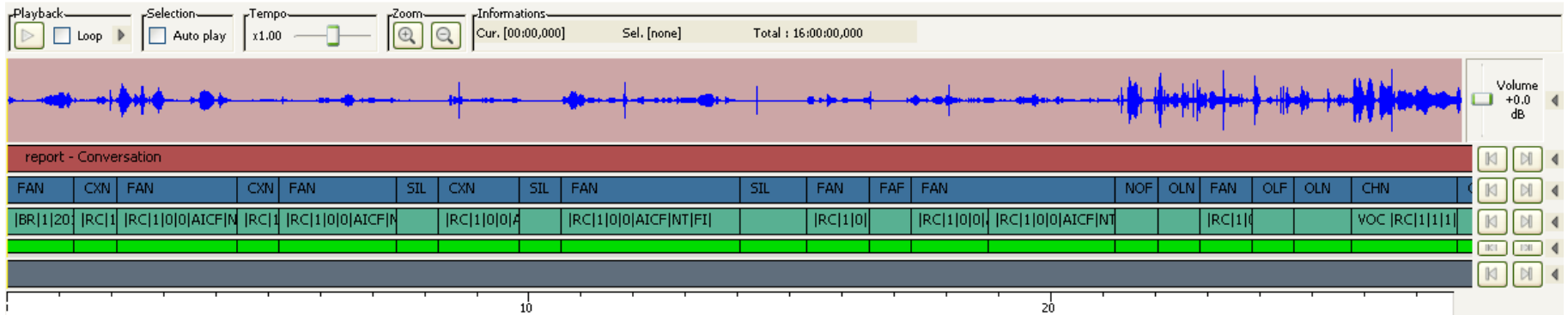


Le système LENA[®]

Fonctionnalités de LENA

3. Traitement des données

- 3h à 16h
- Pré-segmentation visualisable « Transcriber »





Le système LENA[®]

Fonctionnalités de LENA

3. Traitement des données

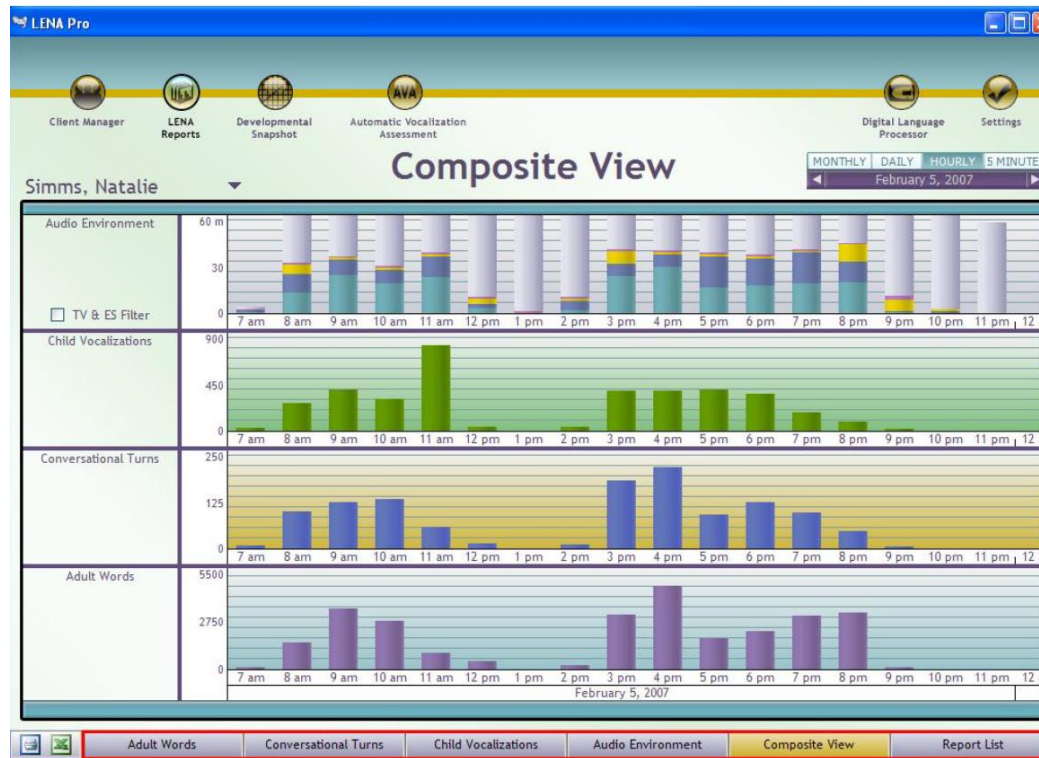
- Rapports LENA (sous forme de graphiques et chiffres) : mots de l'adulte, vocalisations de l'enfant, tours de parole, environnement audio.
- Rapports visualisables : mois, jour, heure, 5 minutes



Le système LENA[®]

Fonctionnalités de LENA

3. Traitement des données





Le système LENA[®]

Fonctionnalités de LENA

3. Traitement des données

Les mots de l'adulte (Adult Word Count : AWC)

Productions de l'adulte compréhensibles, produites à proximité de l'enfant (3m) : chevauchement, voix faible, bruits et sons TV et radio sont exclus.



Le système LENA[®]

Fonctionnalités de LENA

3. Traitement des données

Vocalisations de l'enfant (CV)

Une vocalisation est une séquence de parole continue entourée ou interrompue par une pause de 300 ms ou interrompue par un changement de locuteur.

Ex : [babababa] : 1 vocalisation
[baba#baba] : 2 vocalisations

Phrases, mots, séquences babillées, sons de communication pré-linguistique mais les cris, rires, sons végétatifs sont exclus.



Le système LENA[®]

Fonctionnalités de LENA

3. Traitement des données

Tours de parole

Nombre d'interactions engageant l'enfant et l'adulte.

Vocalisation de l'enfant suivie d'une réponse de l'adulte dans les 5 secondes ou inversement : chevauchements de discours, cris et sons végétatifs sont exclus.



Le système LENA[®]

Fonctionnalités de LENA

3. Traitement des données

Environnement audio

Composantes audio de l'environnement acoustique : TV et radio, parole, discours éloigné, silence, le bruit.



Le système LENA[®]

Congruence Homme - Machine

Un système conçu pour l'anglais (Xu et al. 2009) : 70h d'enregistrement transcrites

Table 3: LENA System Sensitivity: Segmentation agreement between human transcribers and LENA software V3.1.0.

		LENA System			
		Adult	Child	TV	Other
Human Transcribers	Adult	82%	2%	4%	12%
	Child	7%	76%	0%	17%
	TV	8%	0%	71%	21%
	Other	14%	4%	6%	76%



Le système LENA[®]

Congruence Homme - Machine

Un système conçu pour l'anglais (Xu et al. 2009) : 70h d'enregistrement transcrites

Table 3: LENA System Sensitivity: Segmentation agreement between human transcribers and LENA software V3.1.0.

		LENA System			
		Adult	Child	TV	Other
Human Transcribers	Adult	82%	2%	4%	12%
	Child	7%	76%	0%	17%
	TV	8%	0%	74%	18%
	Other	14%	0%	0%	86%

Table 4: Human and LENA-algorithmic based Detection and Classification of Sound as Either Speech or Non-Speech Key Child Vocalizations.

		LENA System	
		Child Vocalizations	Child Cry/Veg/Fixed
Human Transcribers	Child Vocalizations	75%	25%
	Child Cry/Veg/Fixed	16%	84%



Le système LENA[®]

Congruence Homme - Machine

Un système conçu pour l'anglais (Xu et al. 2009) : 70h d'enregistrement transcrites

Table 3: LENA System Sensitivity: Segmentation agreement between human transcribers and LENA software V3.1.0.

		LENA System			
		Adult	Child	TV	Other
Human Transcribers	Adult	82%	2%	4%	12%
	Child	7%	76%		
	TV	8%	0%		
	Other	14%	4%		

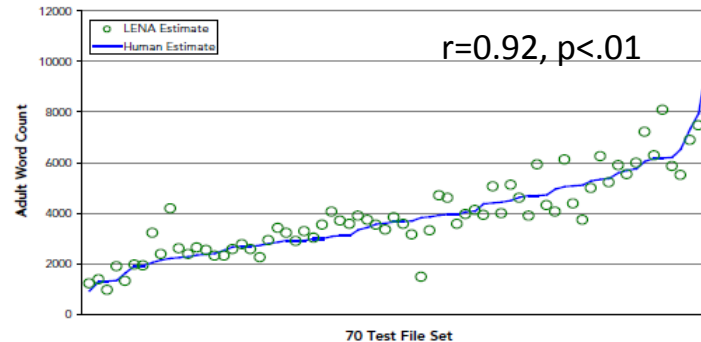


Figure 1. Human and LENA-based AWC estimates for 70 test files.



Le système LENA[®]

Congruence Homme - Machine

Systeme validé pour :

- ✓ Anglais américain (Xu *et al.*, 2009 ; Oetting *et al.*, 2009 ; Gilkerson & Richards, 2008)
- ✓ Anglais canadien (VanDam, M., & Silbert, N. H., 2013 ; Soderstrom & Wittebottle, 2013)
- ✓ Mandarin (Zhang, 2013 ; Gilkerson *et al.*, 2015)
- ✓ Espagnol (Weisleder & Fernald, 2013)
- ✓ Français (Canault *et al.*, 2016)
- ✓ Vietnamien (Ganek & Eriks-Brophy, 2013 ; Ganek *et al.*, 2017)

En cours :

- ✓ Coréen (Pae, 2013)
- ✓ Arabe (Aldosari *et al.*, 2012)
- ✓ Néerlandais (Berends, 2015)



Le système LENA[®]

Congruence Homme - Machine

Variabilité des échantillons

Auteurs	Langue	Echantillon	Age	LENA/Humain	Variables observées	r	% d'accord
Xu, D., Yapanel, U. & Gray, S. 2009	Anglais américain	N= 70	2-36 months	Six portions de 10-minutes d'enregistrements (1h) 70h	AWC CVC	.92	75%
Oetting et al., 2009	Anglais américain	N= 17	2-5 years	Portions de 30-minutes 8.5 h	AWC CTC	.71 - .85 .08-.14	
VanDam, M., & Silbert, N. H., 2013	Anglais canadien	N =26	2 years	2340 segments (30 MAN, FAN et KEY)	CHN FAN MAN		85.9% 59.6% 60.8%
Soderstrom, M. & Wittebolle, K., 2013	Anglais canadien	N=11	12-29 months	183 portions de 5 minutes 15h	AWC CVC	.764 .688	
Weisleder & Fernald, 2013	Espagnol	N= 10	19-24 months	Six portions de 10-minutes d'enregistrements (1h)	AWC	.80	
Gilkerson et al., 2015	Mandarin	N= 22 N=19 10 boys 20 girls	3-23 months	Portions de 15-minutes d'enregistrement 5.5 h	AWC CT	.73 .22 - .72	
Canault et al. , 2016	Français	N= 18 9 boys 9 girls	3-48 months	Six portions de 10-minutes d'enregistrement (1h par jour) 54h	AWC CVC	.64 .71	
Ganek et al., 2017	Vietnamien	N=10	22-42 mois	1h40min	CT	.70	



Le système LENA[®]

Congruence Homme - Machine

Variabilité des échantillons

Auteurs	Langue	Echantillon	Age	LENA/Humain	Variables observées	r	% d'accord
Xu, D., Yapanel, U. & Gray, S. 2009	Anglais américain	N= 70	2-36 months	Six portions de 10-minutes d'enregistrements (1h) 70h	AWC CVC	.92	75%
Oetting et al., 2009	Anglais américain	N= 17	2-5 years	Portions de 30-minutes 8.5 h	AWC CTC	.71 - .85 .08-.14	
VanDam, M., & Silbert, N. H., 2013	Anglais canadien	N =26	2 years	2340 segments (30 MAN, FAN et KEY)	CHN FAN MAN		85.9% 59.6% 60.8%
Soderstrom, M. & Wittebolle, K., 2013	Anglais canadien	N=11	12-29 months	183 portions de 5 minutes 15h	AWC CVC	.764 .688	
Weisleder & Fernald, 2013	Espagnol	N= 10	19-24 months	Six portions de 10-minutes d'enregistrements (1h)	AWC	.80	
Gilkerson et al., 2015	Mandarin	N= 22 N=19 10 boys 20 girls	3-23 months	Portions de 15-minutes d'enregistrement 5.5 h	AWC CT	.73 .22 - .72	
Canault et al. , 2016	Français	N= 18 9 boys 9 girls	3-48 months	Six portions de 10-minutes d'enregistrement (1h par jour) 54h	AWC CVC	.64 .71	
Ganek et al., 2017	Vietnamien	N=10	22-42 mois	1h40min	CT	.70	



Le système LENA[®]

Congruence Homme - Machine

et des résultats

Auteurs	Langue	Echantillon	Age	LENA/Humain	Variables observées	r	% d'accord
Xu, D., Yapanel, U. & Gray, S. 2009	Anglais américain	N= 70	2-36 months	Six portions de 10-minutes d'enregistrements (1h) 70h	AWC CVC	.92	75%
Oetting et al., 2009	Anglais américain	N= 17	2-5 years	Portions de 30-minutes 8.5 h	AWC CTC	.71 - .85 .08-.14	
VanDam, M., & Silbert, N. H., 2013	Anglais canadien	N =26	2 years	2340 segments (30 MAN, FAN et KEY)	CHN FAN MAN		85.9% 59.6% 60.8%
Soderstrom, M. & Wittebolle, K., 2013	Anglais canadien	N=11	12-29 months	183 portions de 5 minutes 15h	AWC CVC	.764 .688	
Weisleder & Fernald, 2013	Espagnol	N= 10	19-24 months	Six portions de 10-minutes d'enregistrements (1h)	AWC	.80	
Gilkerson et al., 2015	Mandarin	N= 22 N=19 10 boys 20 girls	3-23 months	Portions de 15-minutes d'enregistrement 5.5 h	AWC CT	.73 .22 - .72	
Canault et al. , 2016	Français	N= 18 9 boys 9 girls	3-48 months	Six portions de 10-minutes d'enregistrement (1h par jour) 54h	AWC CVC	.64 .71	
Ganek et al., 2017	Vietnamien	N=10	22-42 mois	1h40min	CT	.70	



Le système LENA[®]

Congruence Homme - Machine

et des résultats

Auteurs	Langue	Echantillon	Age	LENA/Humain	Variables observées	r	% d'accord
Xu, D., Yapanel, U. & Gray, S. 2009	Anglais américain	N= 70	2-36 months	Six portions de 10-minutes d'enregistrements (1h) 70h	AWC CVC	.92	75%
Oetting et al., 2009	Anglais américain	N= 17	2-5 years	Portions de 30-minutes 8.5 h	AWC CTC	.71 - .85 .08-.14	
VanDam, M., & Silbert, N. H., 2013	Anglais canadien	N =26	2 years	2340 segments (30 MAN, FAN et KEY)	CHN FAN MAN		85.9% 59.6% 60.8%
Soderstrom, M. & Wittebolle, K., 2013	Anglais canadien	N=11	12-29 months	183 portions de 5 minutes 15h	AWC CVC	.764 .688	
Weisleder & Fernald, 2013	Espagnol	N= 10	19-24 months	Six portions de 10-minutes d'enregistrements (1h)	AWC	.80	
Gilkerson et al., 2015	Mandarin	N= 22 N=19 10 boys 20 girls	3-23 months	Portions de 15-minutes d'enregistrement 5.5 h	AWC CT	.73 .22 - .72	
Canault et al. , 2016	Français	N= 18 9 boys 9 girls	3-48 months	Six portions de 10-minutes d'enregistrement (1h par jour) 54h	AWC CVC	.64 .71	
Ganek et al., 2017	Vietnamien	N=10	22-42 mois	1h40min	CT	.70	



Le système LENA[®]

Congruence Homme - Machine

...et des résultats

Auteurs	Langue	Echantillon	Age	LENA/Humain	Variables observées	r	% d'accord
Xu, D., Yapanel, U. & Gray, S. 2009	Anglais américain	N= 70	2-36 months	Six portions de 10-minutes d'enregistrements (1h) 70h	AWC CVC	.92	75%
Oetting et al., 2009	Anglais américain	N= 17	2-5 years	Portions de 30-minutes 8.5 h	AWC CTC	.71 - .85 .08-.14	
VanDam, M., & Silbert, N. H., 2013	Anglais canadien	N =26	2 years	2340 segments (30 MAN, FAN et KEY)	CHN FAN MAN		85.9% 59.6% 60.8%
Soderstrom, M. & Wittebolle, K., 2013	Anglais canadien	N=11	12-29 months	183 portions de 5 minutes 15h	AWC CVC	.764 .688	
Weisleder & Fernald, 2013	Espagnol	N= 10	19-24 months	Six portions de 10-minutes d'enregistrements (1h)	AWC	.80	
Gilkerson et al., 2015	Mandarin	N= 22 N=19 10 boys 20 girls	3-23 months	Portions de 15-minutes d'enregistrement 5.5 h	AWC CT	.73 .22 - .72	
Canault et al. , 2016	Français	N= 18 9 boys 9 girls	3-48 months	Six portions de 10-minutes d'enregistrement (1h par jour) 54h	AWC CVC	.64 .71	
Ganek et al., 2017	Vietnamien	N=10	22-42 mois	1h40min	CT	.70	



Le système LENA[®]

Focus sur le français (Canault *et al.*, 2016)

Behav Res (2016) 48:1109–1124
DOI 10.3758/s13428-015-0634-8



Reliability of the Language ENvironment Analysis system (LENA[™]) in European French

Mélanie Canault¹ · Marie-Thérèse Le Normand^{2,6} · Samy Foudil³ · Natalie Loundon⁴ ·
Hung Thai-Van⁵

18 enfants

6 tranches d'âge (0-6 mois ; 7-12 mois ; 13-18 mois ; 19-24 mois ; 25-36 mois ; 37-48 mois)

3 jours en 1 semaine (10 à 16h/enregistrement)

3h par enfant (transcription de 6 portions de 10 min / enregistrement)

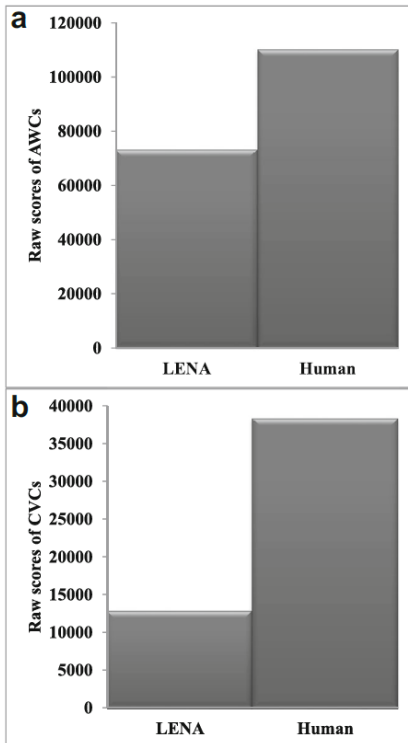
Variables : AWC et CV

Total : 54h intégralement transcrites



Le système LENA[®]

Focus sur le français (Canault *et al.*, 2016)



Comptages humains > Comptages LENA

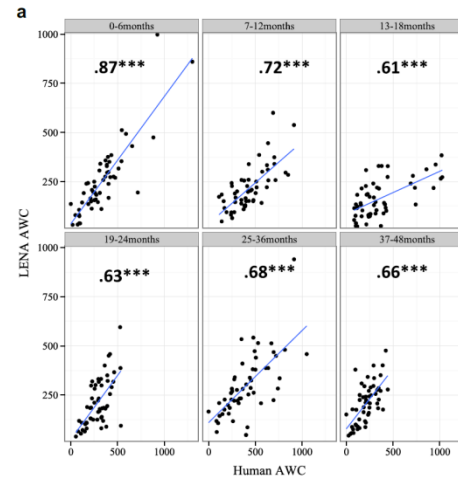
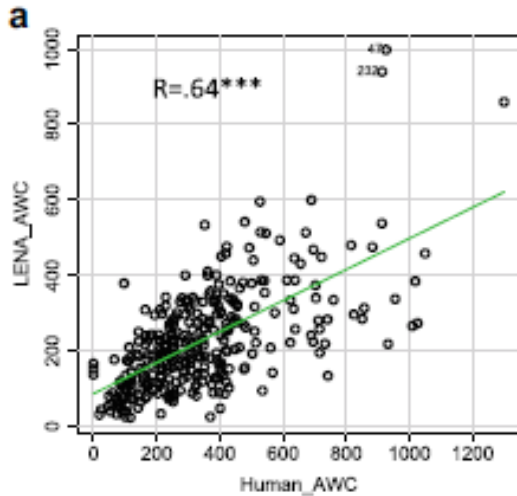
AWC : 110.318 (Homme) vs. 73.274 (LENA) (ratio of 1.56).

CVC : 38.409 (Homme) vs. 12.881 (LENA) (ratio of 2.86).



Le système LENA[®]

Focus sur le français (Canault *et al.*, 2016)

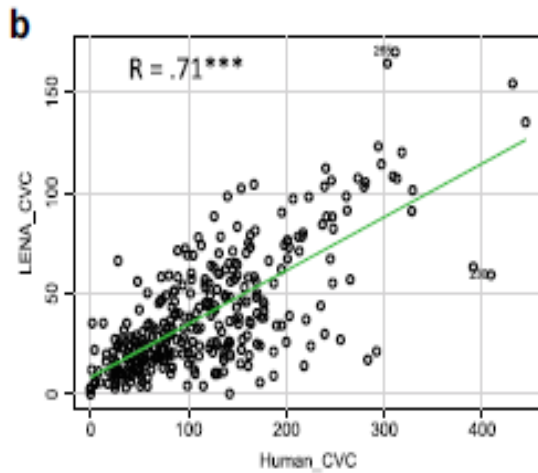


***p < .001

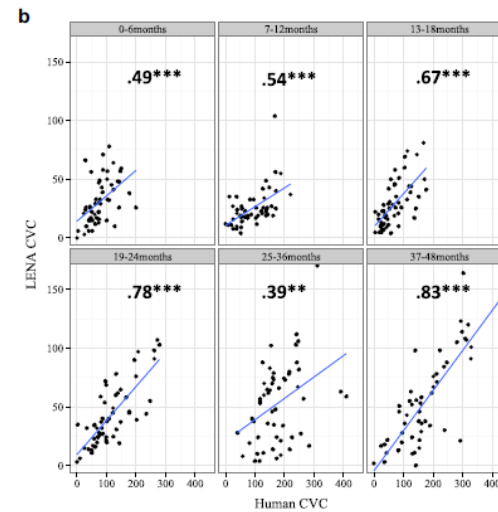


Le système LENA[®]

Focus sur le français (Canault *et al.*, 2016)



*** $p < .001$





Le système LENA[®]

Focus sur le français (Canault *et al.*, 2016)

En première intention,

1. On ne peut pas considérer les chiffres pour eux-mêmes

Mais,

2. Les corrélations sont bonnes : LENA intéressant pour les suivis longitudinaux par exemple.



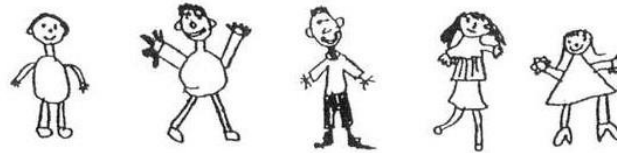
Les nouvelles pistes

Intérêt pour la recherche fondamentale et la pratique clinique

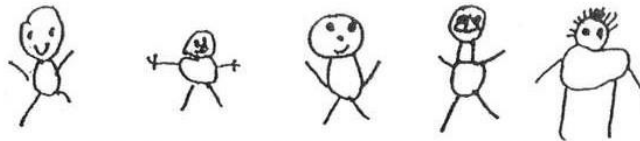
- Exploration du Langage Adressé à l'Enfant ou produit par l'enfant chez des populations atypiques (autistes, IC, prématurés...)
- Programme d'intervention *Thirty Million Words Initiative* : sensibilisation des familles à risque à l'influence de l'*input* afin d'améliorer l'interaction parent-enfant (Suskind, Leffel, Hernandez, Sapolich, Suskind, Kirkham, et al., 2016 ; Zhang, Xu, Jiang, Gilkerson, Xu, Richards & Topping, 2015)



Merci de votre attention !!!



*Dessins d'enfants de 5 à 6 ans exposés à moins
de une heure de télévision par jour.*



*Dessins d'enfants de 5 à 6 ans exposés à plus
de trois heures de télévision par jour.*

SOURCE : TV LOBOTOMIE. MICHEL DESMURGET. P. 136. ED. MAX MILO.



Bibliographie

- Aldosari, M., Almuslanani, A., Wilson, F. & Gilkerson, J. (2012, May). Validation of Language Environment Analysis (LENA) in Arabic-speaking individuals. Poster session presented at the International Meeting for Autism Research, Toronto, Ontario, Canada.
- Berends, C. (2015). The LENA System in Parent-Child Interaction in Dutch Preschool Children with Language Delay [M.A. thesis]. Utrecht, Holland:UMC-Utrecht; 2015.
- Canault, M., Le Normand, M.-T., Foudil, S., Loundon, N., & Thai-Van, H. (2016). Reliability of the Language Environment Analysis system (LENA™) in European French. *Behavior Research Methods*, 48(3), 1109-1124. doi: [10.3758/s13428-015-0634-8](https://doi.org/10.3758/s13428-015-0634-8)
- Ganek, H. & Eriks-Brophy, A. (2013, April). Questions in using LENA with Vietnamese and Canadian children with hearing loss. Poster session presented at the LENA International Conference 2013, Denver, CO.
- Ganek, H. V., & Eriks-Brophy, A. (2017). A Concise Protocol for the Validation of Language ENvironment Analysis (LENA) Conversational Turn Counts in Vietnamese. *Communication Disorders Quarterly*, 152574011770509. <https://doi.org/10.1177/1525740117705094>
- Gilkerson, J. & Richards, J.A. (2008). *The LENA natural language study*. LENA Foundation Technical Report No. LTR-02-2. Boulder, CO: LENA Research Foundation. Available at: http://www.lenafoundation.org/wp-content/uploads/2014/10/LTR-02_2_Natural_Language_Study.pdf
- Gilkerson, J. & Richards, J. A. (2009) The power of talk: Impact of adult talk, conversational turns, and TV during the critical 0-4 years of child development (Technical Report LTR-01-2). Retrieved from: http://www.lenababy.com/pdf/The_Power_of_Talk.pdf
- Gilkerson, J., Zhang, Y., Xu, D., Richards, J. A., Xu, X., Jiang, F., Harnsberger, J., & Topping, K. (2015). Evaluating LENA System Performance for Chinese: A Pilot Study in Shanghai. *Journal of Speech Language and Hearing Research*. doi:10.1044/2015_JSLHR-L-14-0014
- Jackson, C. W. (2013). Spanish-English speaking preschoolers: An exploratory study examining LENA data. Paper presented at the LENA International Conference 2013, Denver, CO.
- Hart B. & Risley T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes.
- Korman M. (1984). Adaptive aspects of maternal vocalizations in differing contexts at ten weeks. *First Language*, 5, 44-45.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.
- Oetting, J. B., Hartfield, L.R., & Pruitt, S. L. (2009). Exploring LENA as a tool for researchers and clinicians. *The ASHA Leader*. 14(6): 20-22.
- Pae, S. (2013). An investigation of the language environment of young Korean children and the impact of feedback on language development. Paper presented at: LENA International Conference, Denver, CO, April 29, 2013
- Rose, Yvan & Brian MacWhinney. 2014. The PhonBank Project: Data and Software-Assisted Methods for the Study of Phonology and Phonological Development. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (eds.), *The Oxford Handbook of Corpus Phonology*, 380–401. Oxford: Oxford University Press.
- Rose, Yvan. 2012. Multilingual Phonological Corpus Analysis: The Tools behind the PhonBank Project. In Thomas Schmidt & Kai Wörner (eds.), *Multilingual Corpora and Multilingual Corpus Analysis*, 365–381. (Hamburg Studies on Multilingualism). Amsterdam: John Benjamins Publishing Company. <https://benjamins.com/catalog/hsm.14.27ros>.
- Soderstrom, M & Wittebolle, K (2013) When Do Caregivers Talk? The Influences of Activity and Time of Day on Caregiver Speech and Child Vocalizations in Two Childcare Environments. *PLoS ONE* 8(11): e80646. doi:10.1371/journal.pone.0080646
- Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., Levine, S. C. (2016). A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study. *Journal of Child Language*. 43(2), 366-406. doi: 10.1017/S0305000915000033
- VanDam, M., & Silbert, N. H. (2013). Precision and error of automatic speech recognition. *The Journal of the Acoustical Society of America*, 133(5), 3245. <https://doi.org/10.1121/1.4805200>
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143-2152. doi :10.1177/056797613488145
- Wells, G. (1979). Describing children's linguistic development at home and at school. *British Educational Research Journal*, 5(1), 75–98.
- Xu D., Yapanel U. & Gray S. (2009). *Reliability of the LENA Language Environment Analysis system in Young Children's Natural Home Environment*. Boulder, CO: LENA Foundation. *Technical Report LTR-05-2*.
- Zhang Y. (2013). The correlation of family responsiveness with language and cognitive development in infants and toddlers. Paper presented at: LENA International Conference; Denver, CO, April 29, 2013.
- Zhang, Y., Xu, X., Jiang, F., Gilkerson, J., Xu, D., Richards, J. A. & Topping, K. J. (2015). Effects of quantitative linguistic feedback to caregivers of young children: A pilot study in China. *Communication Disorders Quarterly*, 37(1), 16-24. doi: 10.1177/1525740115575771