FREQUENCY AND THE LEXICAL ORGANIZATION OF VERBS

Daniel Cerato Germann, Aline Villavicencio, Maity Siqueira

Federal University of Rio Grande do Sul (Brazil)

Goals

- To investigate the influence of linguistic factors on lexical acquisition and organization
- To build linguistic profiles to help diagnose pathologies that affect language

Overview

- Related Work
- Materials and Methods
- Results
- Conclusions and Future Work

Related Work

- Some factors that may affect language acquisition:
  - Frequency
  - Polysemy
  - Syntactic structure
  - Concreteness
  - Specificity
  - Conventionality
  - Figurativeness

Related Work

- Goldberg's (1999):
  - crosslinguistically polysemic and frequent verbs acquired early
- Korhonen and Buttery (2007):
  - verb type preferences influence acquisition of subcategorization frames
- Tonietto et al (2008)
  - influence of pragmatic aspects on lexical organization of verbs
- **Graph theory and network analysis methods useful in mapping differences in associative structures across groups.**
- Steyvers and Tenenbaum (2005):
  - model of semantic network growth compatible with effects of learning history variables (age of acquisition and frequency) in semantic processing tasks

Goals

- To investigate the influence of linguistic factors on lexical acquisition and organization
- To build linguistic profiles to help diagnose pathologies that affect language

Materials and Methods

- Longitudinal study
- Data from psycholinguistic Naming Task:

- 17 videos of division or destruction actions
- Each action: 1 actor and 1 object

Materials and Methods

- 55 Participants per group
  - G1: 2;0 to 3;11 year old (M=3;1)
  - G2 (G1 after 2 years): 4;1 to 6;6 year old (M=5;5)
  - G3: 17;0 to 34;0 (M=21;8)
- Native speakers of Brazilian Portuguese
- Verbs annotated with frequency:
  - Fscore: occurrences in “Florianópolis” corpus (Scliar-Cabral, 1993; MacWhinney, 2000)

Materials and Methods

- Verbs
  - Manually pre-processed to remove noise
    - Unrelated (*I don't know*) or low frequency (<=1) answers

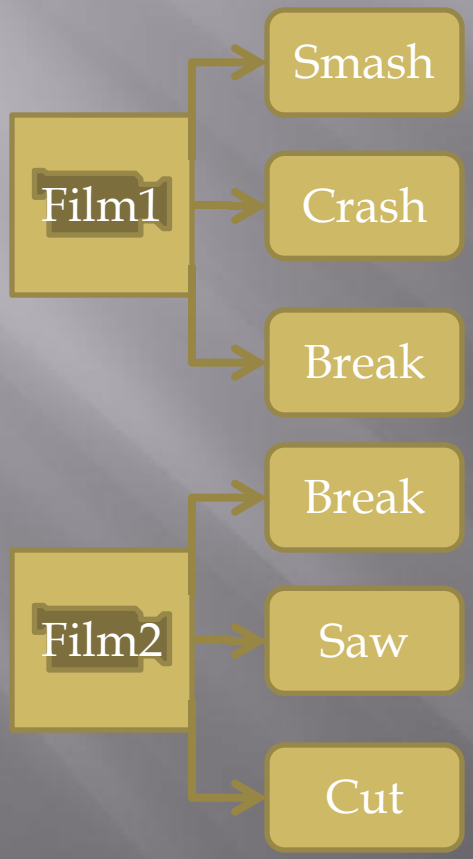
Answers	# per group	# per action	# types
G1	785/935	46.18	6.76
G2	911/935	53.59	5.53
G3	917/935	53.94	4

Materials and Methods

- Verb interactions represented in semantic networks
  - Structural properties of lexicon
    - Network analysis and modification
  - Contents of lexicon
    - Similarity measure

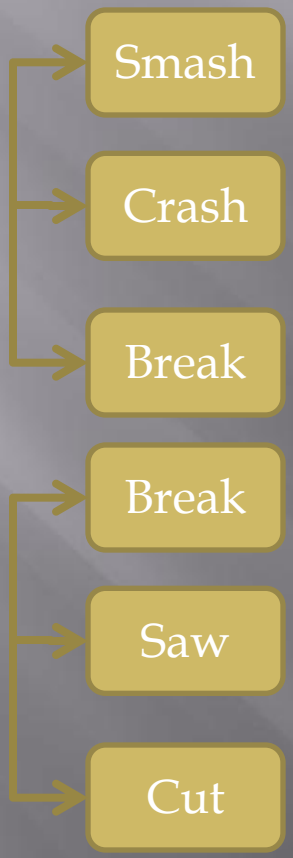
Materials and Methods

- Semantic Network
  - For each film, verbs form a clique



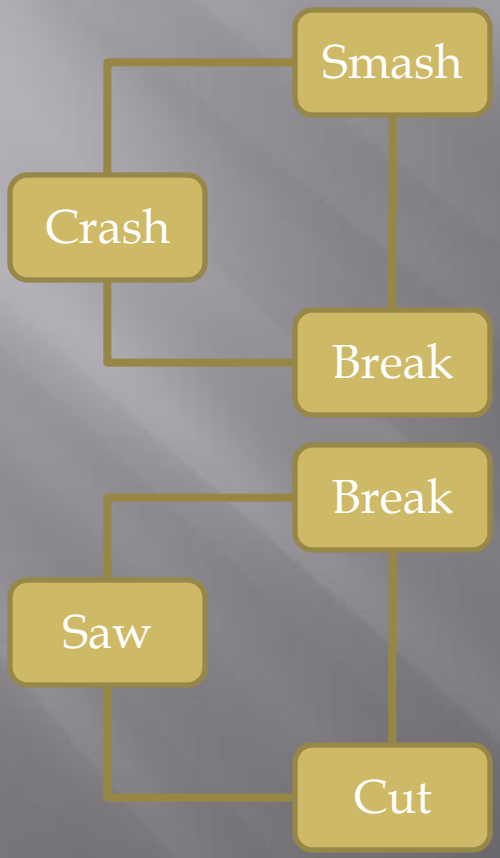
Materials and Methods

- Semantic Network
  - For each film, verbs form a clique



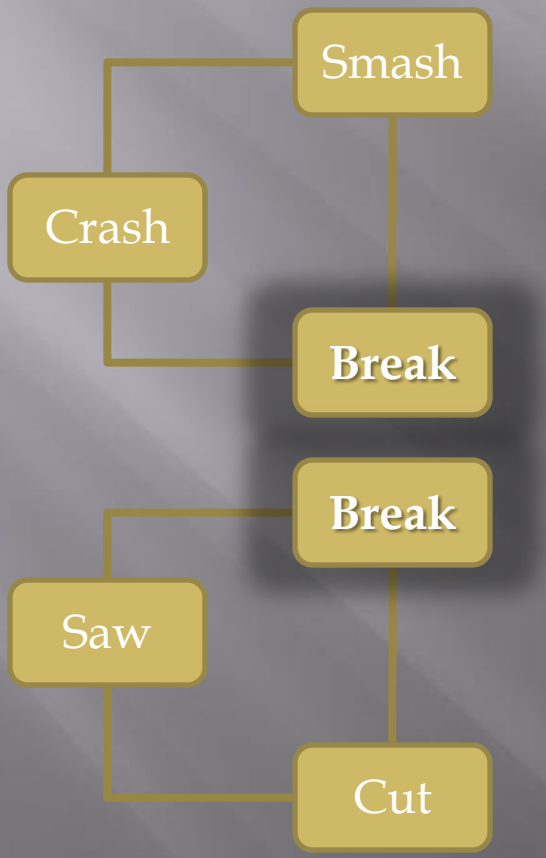
Materials and Methods

- Semantic Network
  - For each film, verbs form a clique



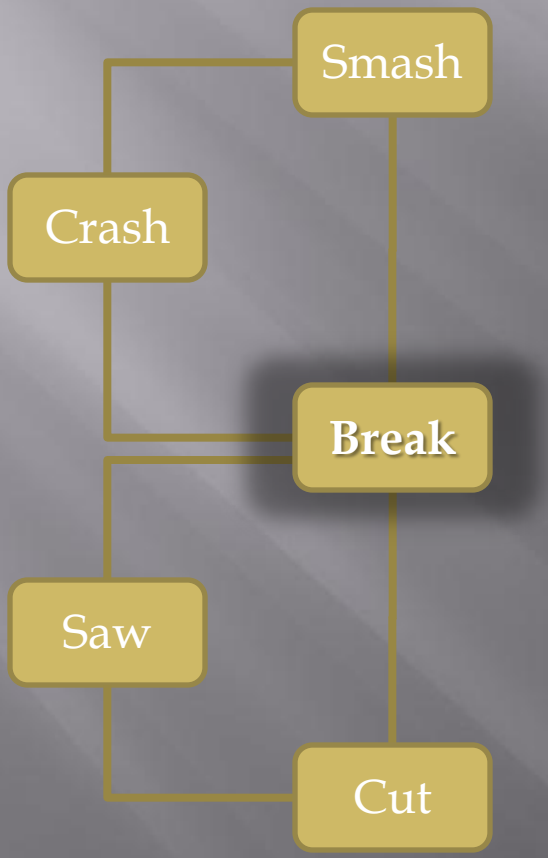
Materials and Methods

- Semantic Network
  - For each film, verbs form a clique

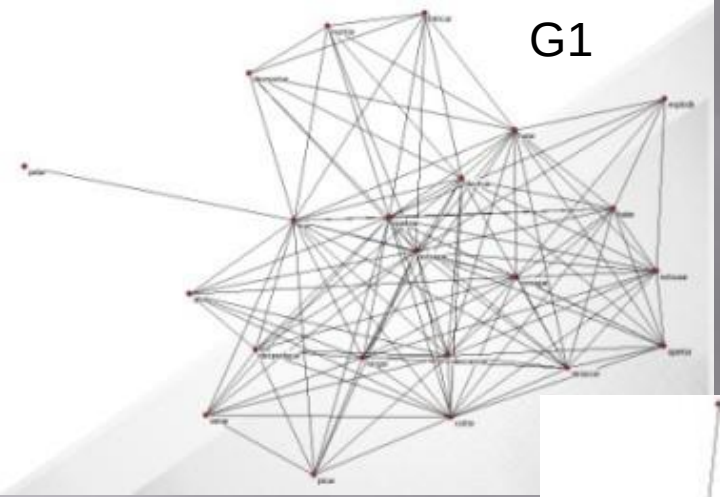


Materials and Methods

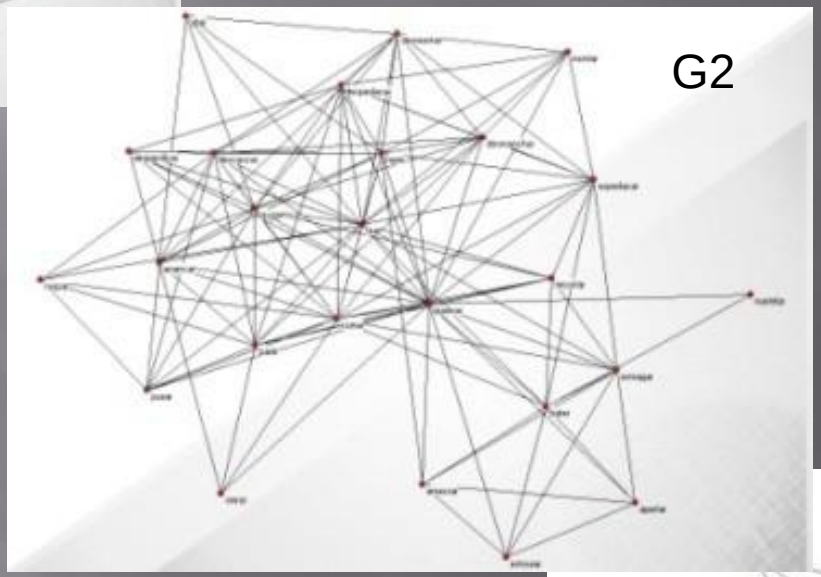
- Semantic Network
  - For each film, verbs form a clique



G1



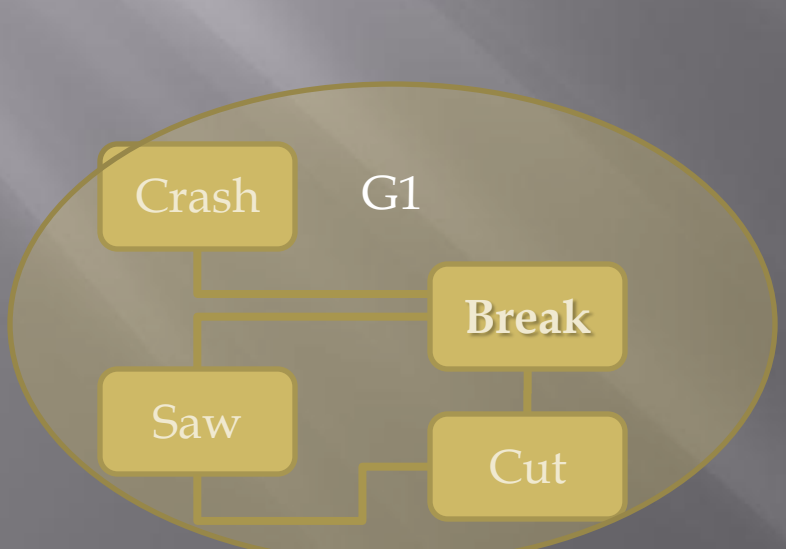
G2

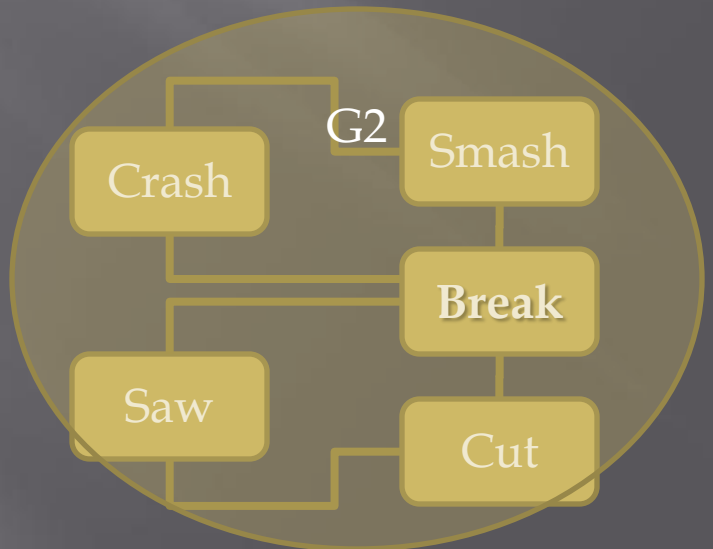


G3

Materials and Methods

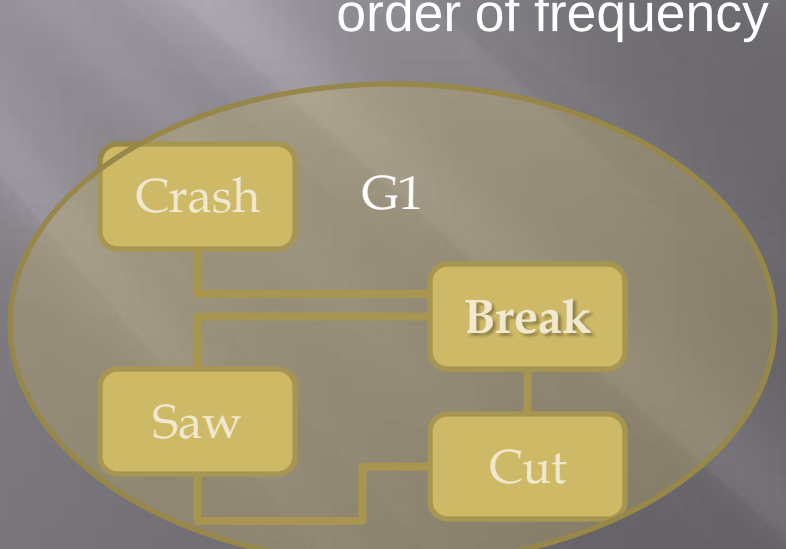
- For each group, one semantic network
- Compare structure and content of groups
  - Will removing less frequent verbs from older group, result in convergence of overall structure to that of a younger group ?
    - **Will we see increase in similarity of content/structure?**
  - **Network modification**

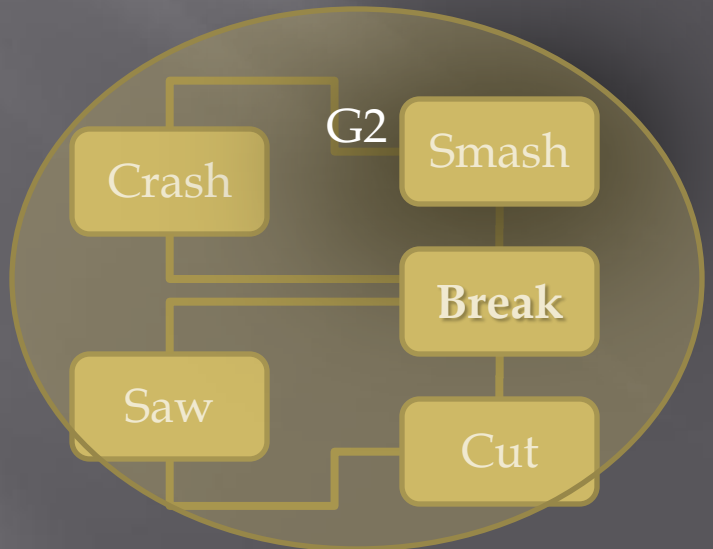




Materials and Methods

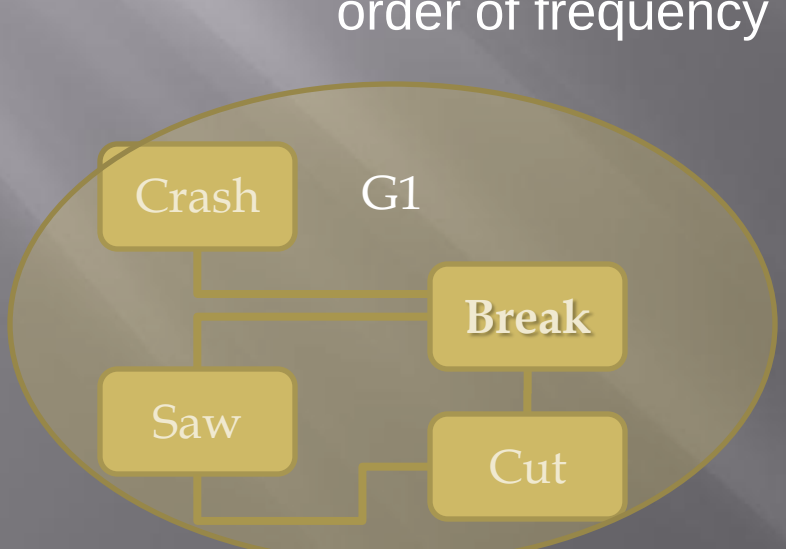
- For each group, one semantic network
- Compare structure and content of groups
  - Will removing less frequent verbs from older group, result in convergence of overall structure to that of a younger group ?
    - **Will we see increase in similarity of content/structure?**
  - **Network modification**
    - Starting with older group and removing verbs in increasing order of frequency

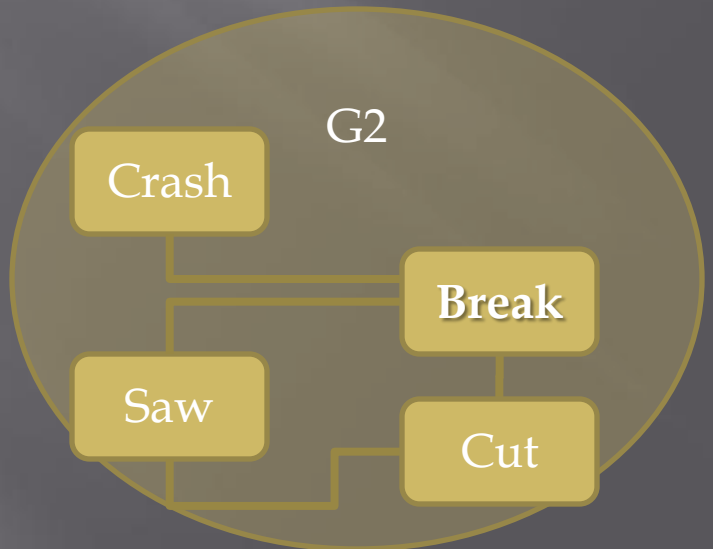




Materials and Methods

- For each group, one semantic network
- Compare structure and content of groups
  - Will removing less frequent verbs from older group, result in convergence of overall structure to that of a younger group ?
    - **Will we see increase in similarity of content/structure?**
  - **Network modification**
    - Starting with older group and removing verbs in increasing order of frequency





Materials and Methods

- Measure similarity through:
  - Jaccard coefficient
  - Number of vertices (*n*)
  - Average minimal path length (*L*)
  - Density (*D*)
  - Number of edges (*M*)
  - Average clustering coefficient (*C/s*)
  - Average node connectivity (*k*),
  - Average number of repetitions (*r*)

Materials and Methods

- Measure similarity through:
  - Jaccard coefficient
  - Number of vertices (*n*)
  - Average minimal path length (*L*)
  - *Density (D)*
  - Number of edges (*M*)
  - Average clustering coefficient (*C/s*)
  - *Average node connectivity (k)*
  - Average number of repetitions (*r*)

Measures of proportion of **# of edges vs # of nodes (semantic share)**

Materials and Methods

- Measure similarity through:
  - Jaccard coefficient
  - Number of vertices (*n*)
  - *Average minimal path length (L)*
  - Density (*D*)
  - Number of edges (*M*)
  - *Average clustering coefficient (C/s)*
  - Average node connectivity (*k*)
  - *Average number of repetitions (r)*

Measures of vocabulary structuring **(small distances, more repetition and high clustering → unstructured lexicon)**

Materials and Methods

- Frequency decreases as age increases:

+

-

Answers	Average type	Average token
G1	44.05	43.44
G2	35.92	35.71
G3	17.84	21.22

Results

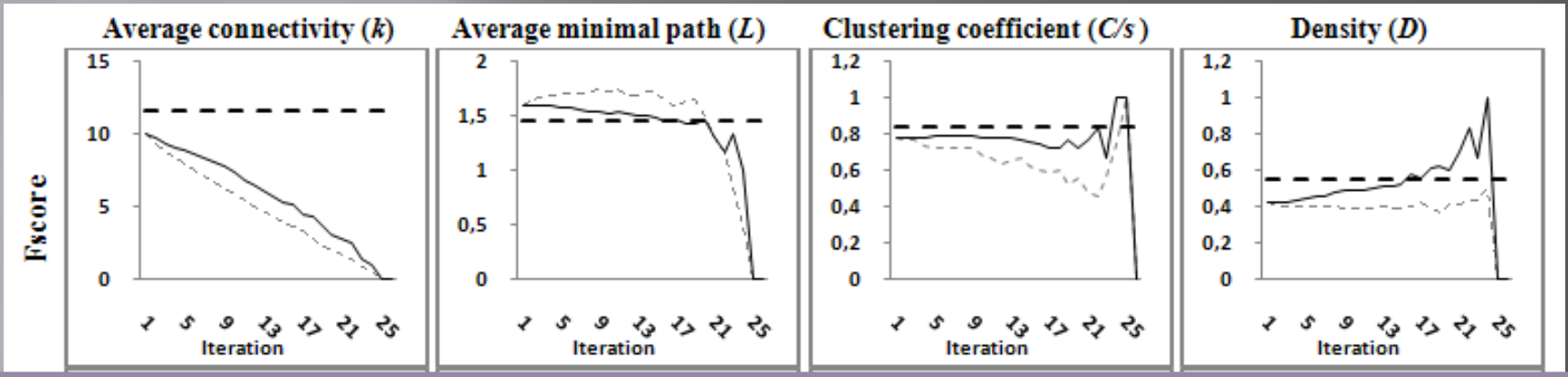
- Increase in vocabulary size and decrease of repetition with age

	G1	G2	G3
Number of vertices (*n*)	22	25	31
Average minimal path length (*L*)	1.46	1.6	1.98
Density (*D*)	0.55	0.42	0.27
Number of edges (*M*)	128	126	126
Average clustering coefficient (*C/s*)	0.84	0.78	0.78
Average node connectivity (*k*),	μ = 11.64, SD = 6.73	μ = 10.08, SD = 4.86	μ = 8.13, SD = 4.76
Average number of repetitions (*r*)	μ = 5.23, SD = 4.41	μ = 3.76, SD = 3.15	μ = 2.19, SD = 1.58

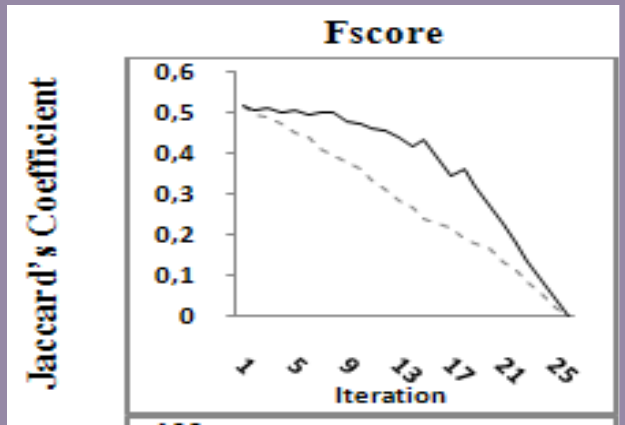
Results

- Networks get structurally closer.

G2-G1



- Networks get closer in content:
  - Jaccard score: G1-G2 = 0.52 G2-G3 = 0.44



Results

- Frequency decreases with age
- Vocabulary grows and specializes with age
- Children’s groups more similar than adults.

Conclusions and Future Work

- Investigation of influence of linguistic factors on lexical acquisition and organization
  - Frequency
    - Different network topology for different groups
    - Different content for different groups
- Build linguistic profiles to help diagnose pathologies that affect language

Conclusions and Future Work

- Test other factors: polyssemy, concreteness, syntactic complexity
- Investigate lexical dissolution in pathologies (Alzheimer's disease)
- Obtain larger data sets
- Test with other categories (nouns, expressions,…)

Acknowledgments

Research partly supported by CNPq (Projects 479824/2009-6 and 309569/2009-5), FINEP and SEBRAE (COMUNICA project FINEP/SEBRAE 1194/07).

Kathryn A. Coronges, Alan W. Stacy, and Thomas W. Valente. 2007. Structural Comparison of Cognitive Associative Networks in Two Populations. *Journal of Applied Social Psychology*, 37(9): 2097-2129

Simon de Deyne and Gert Storms. 2008. Word associations: Network and semantic properties. *Behavior Research Methods*, 40(1): 213-231.

Bento C. Dias da Silva et al. 2000. Construção de um thesaurus eletrônico para o português do Brasil. In *Proceedings of the 4th Processamento Computacional do Português Escrito e Falado (PROPOR)*, 1-10.

Antônio Houaiss. 2007. *Dicionário Eletrônico Houaiss da Língua Portuguesa*, version 2.0a. Editora Objetiva.

Adele E. Goldberg. The Emergence of the Semantics of Argument Structure Constructions. 1999. In *Emergence of Language*. Lawrence Erlbaum Associates , Mahwah, NJ.

Erick Galani Maziero, E.G. et al. 2008. A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Eletrônico para o Português do Brasil. In P*roceedings of the 6th Workshop em Tecnologia da Informação e da Linguagem Humana*.

Mark Steyvers and Joshua B. Tenenbaum. 2005. The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science: A Multidisciplinary Journal*, 29(1): 41-78.

Lauren Tonietto. 2009. *Desenvolvimento da convencionalidade e especificidade na aquisição de verbos: relações com complexidade sintática e categorização*. Ph.D. thesis, Federal University of Rio Grande do Sul.

Lauren Tonietto, Aline Villavicencio, Maity Siqueira, Maria Alice de Mattos Pimenta Parente, Tania Mara Sperb. 2008. A especificidade semântica como fator determinante na aquisição de verbos. *Psico*, 39(3): 343-351.