UNIVERSITÄT
LEIPZIG

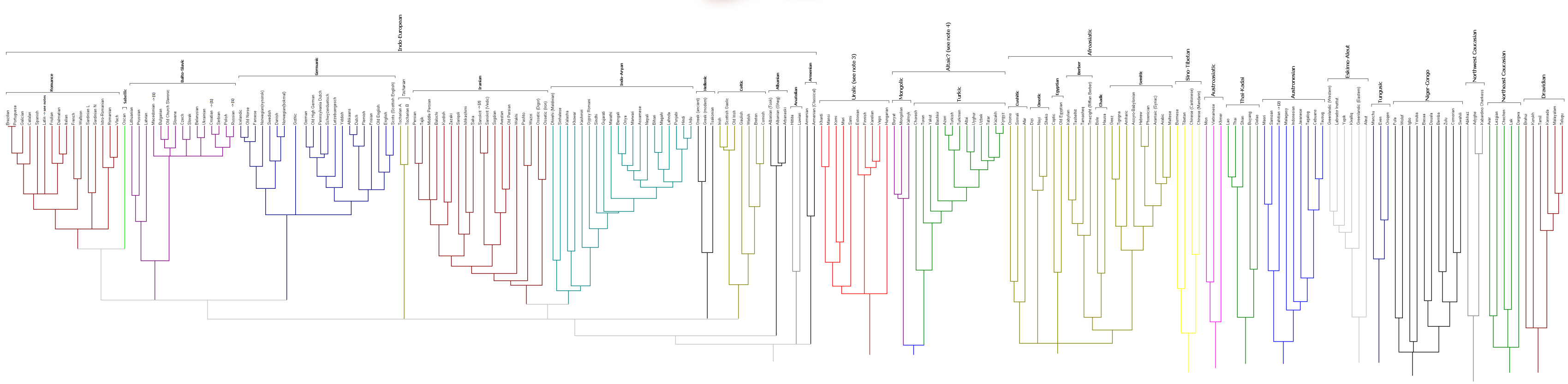
A Dynamic Network Approach to bilingual child data

Antje Quick, Stefan Hartmann,
Nikolas Koch & Paul Ibbotson

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Language Acquisition

> "For languages to survive as complex cultural systems, they need to be learnable." (Ibbotson et al. 2019)



How do children acquire language(s)

Complexity of languages is remarkable - Different research traditions

- ✓ children are equipped with pre-existing experience (e.g. Chomsky 1965)

- ✓ children acquire languages by actively constructing complexity (e.g. Tomasello 2003) → piecemeal acquisition

Language acquisition from a usage-based perspective

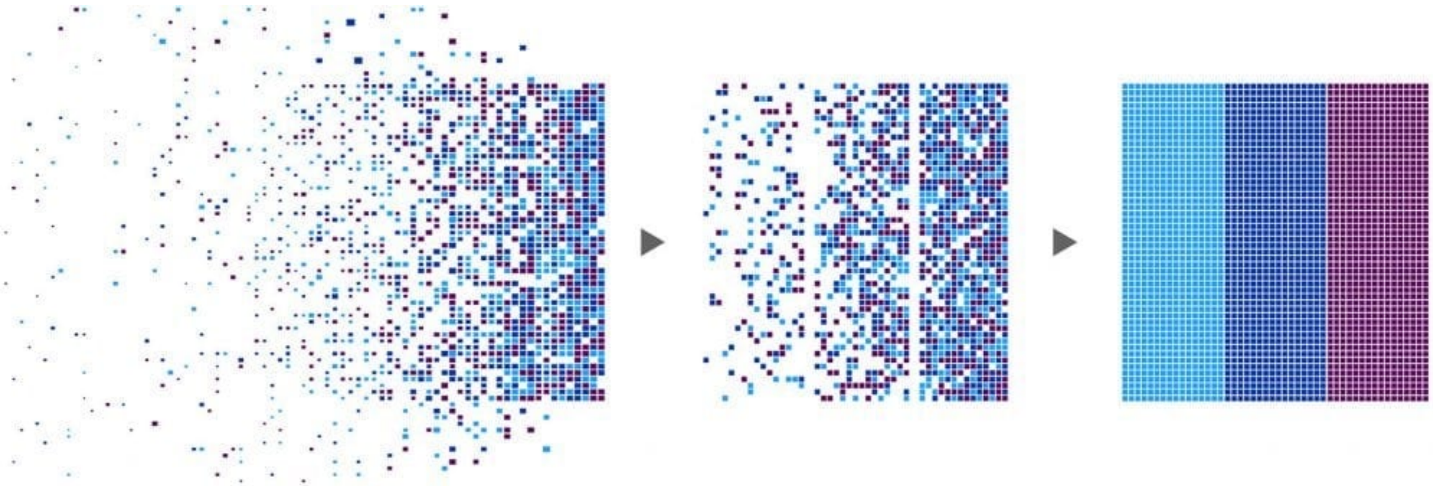
Basic assumptions:

1. Language acquisition is primarily based on the linguistic input that learners receive from their social environment.
2. Drawing upon their social-cognitive abilities, learners derive linguistic patterns from the language input and gradually generalize the rules of the respective language(s).

Tomasello (2009)

Humans are exquisitely adept at finding patterns in language.

Building up language(s)



- Key assumptions are the ability to recognize patterns (Tomasello 2009; *pattern finding*) and,
- the pattern-based nature of the input (e.g., Cameron-Faulkner et al. 2003; Behrens 2006; Christiansen & Chater 2008).

Grafik: https://www.plantvision.se/wp-content/uploads/2018/05/Pattern-Recognition-e1527681213449.jpg

Study

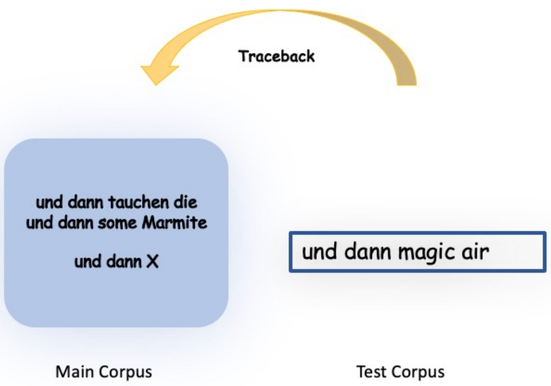
➢ patterns in child-directed speech (CDS) are a significant predictor of patterns used by children (e.g., Ambridge et al., 2015; Lieven, Pine & Baldwin, 1997 Diessel, 2007; Tomasello, 2003)

children are able to extract linguistic knowledge from the input they receive and in a piecemeal way gradually build up complexity

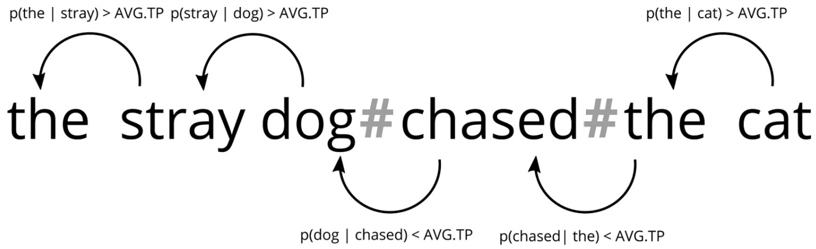
Four stages of construction acquisition according to Tomasello (2003)

How to account for patterns?

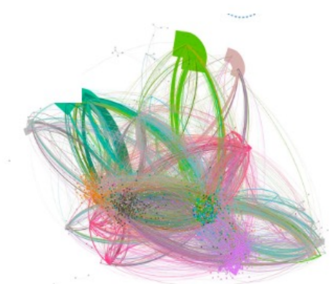
Traceback



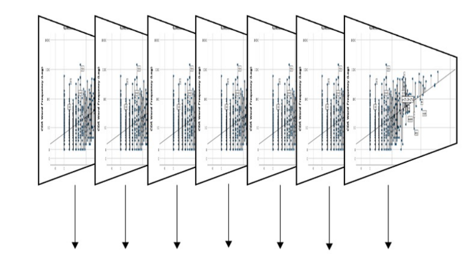
Chunk-based learner



Dynamic Network Model



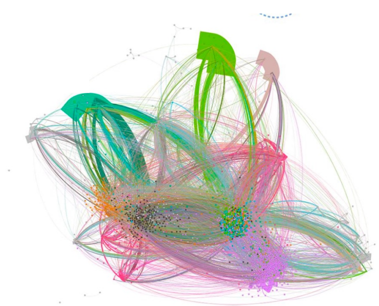
Frequency Filter



Dynamic Network Model

- ✓ Dynamic Network Model (DNM) boils down to a combination of two measures:
  - ✓ word frequencies
  - ✓ transition probabilities

Network builds up patterns of use based on distributional information



I want cake.
I want milk.

I want cake
I want milk

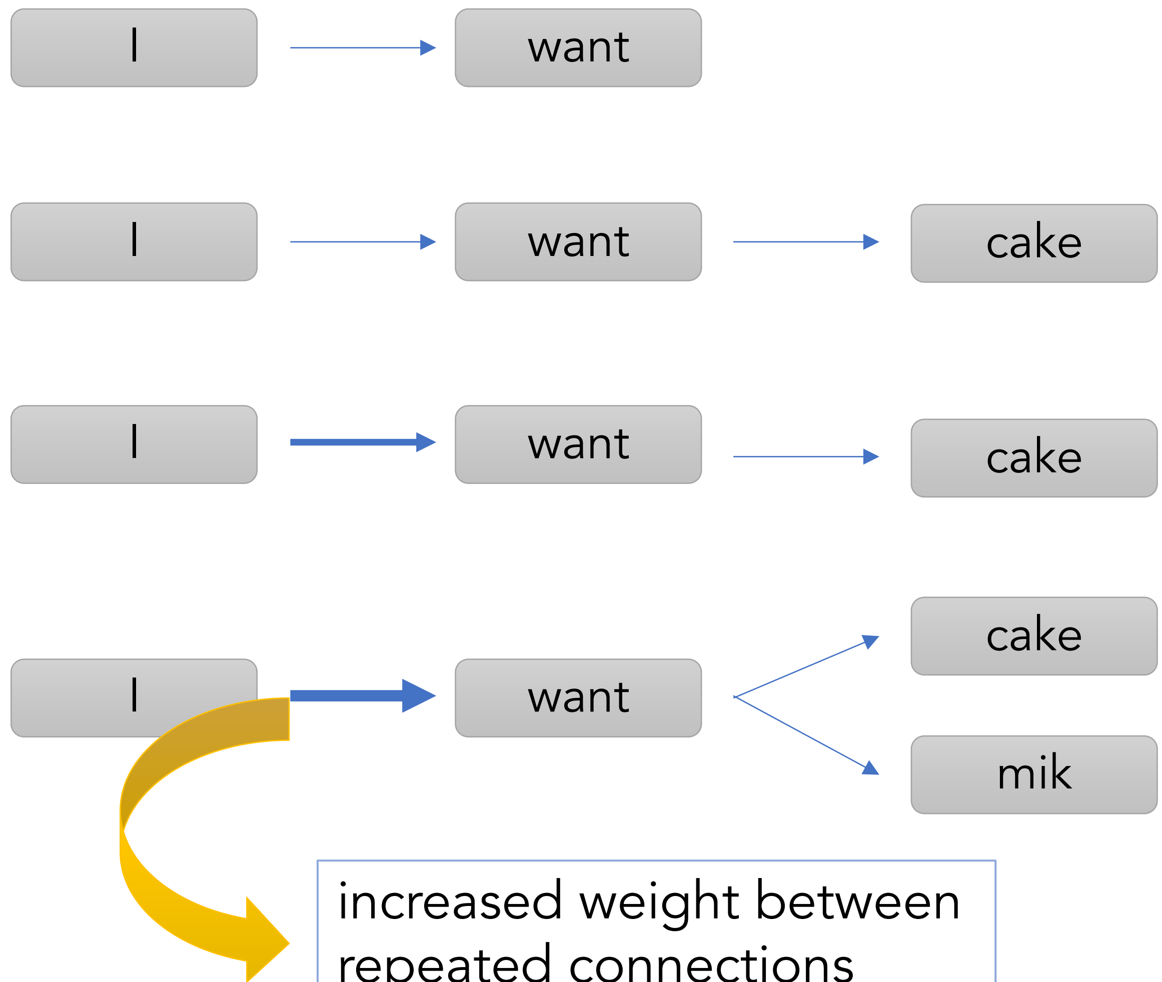
community detection -

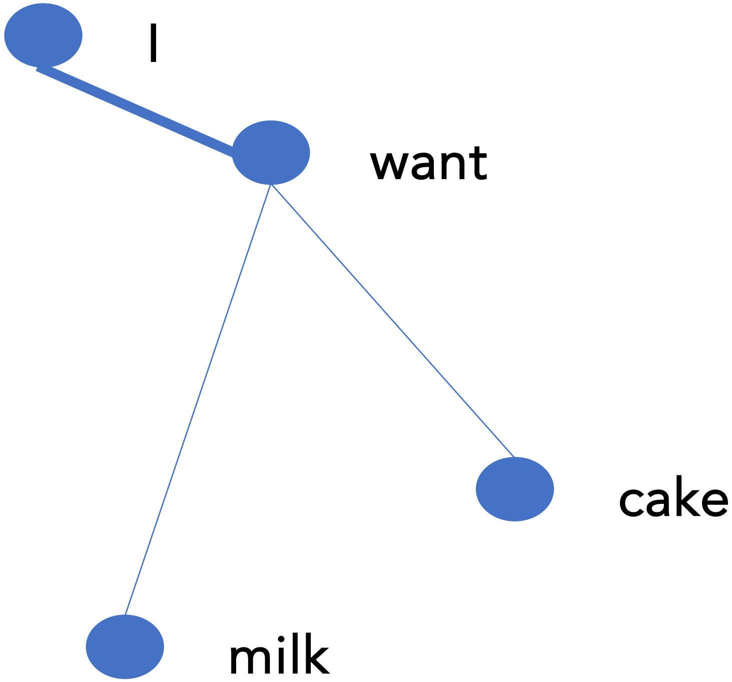
I want

want cake

I want

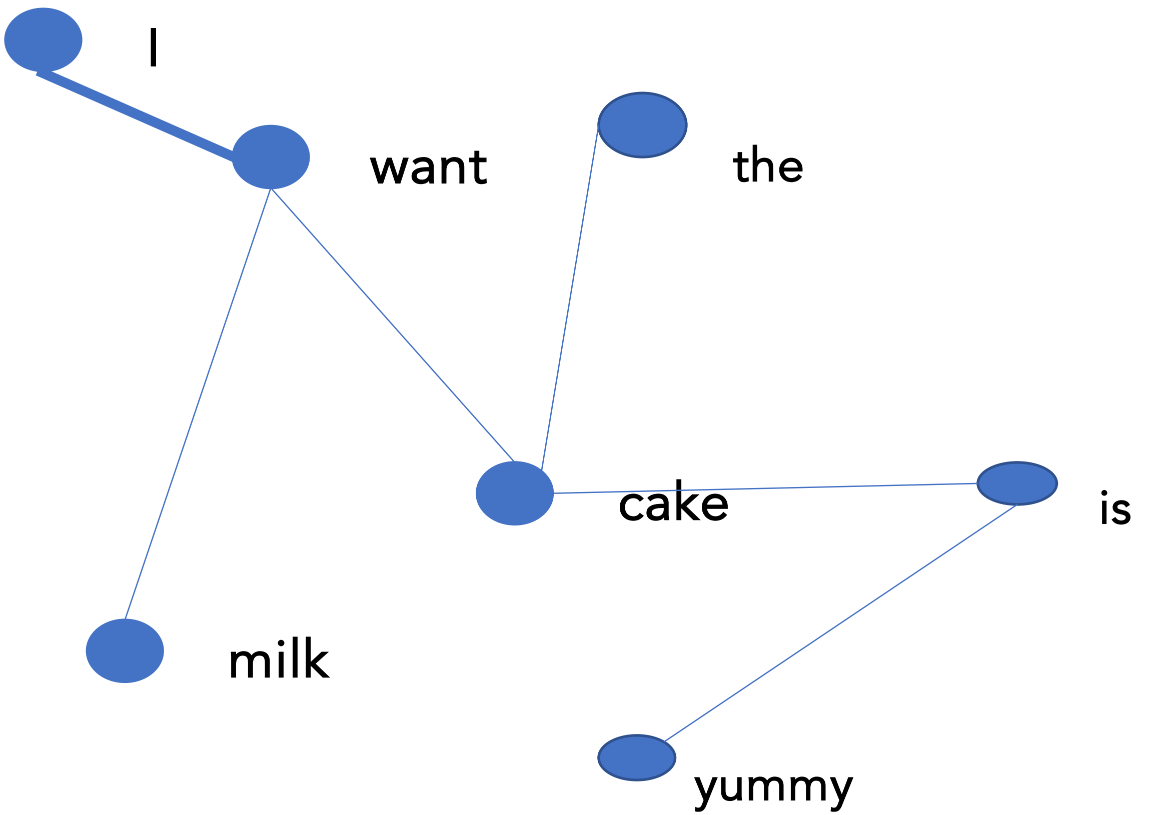
want milk



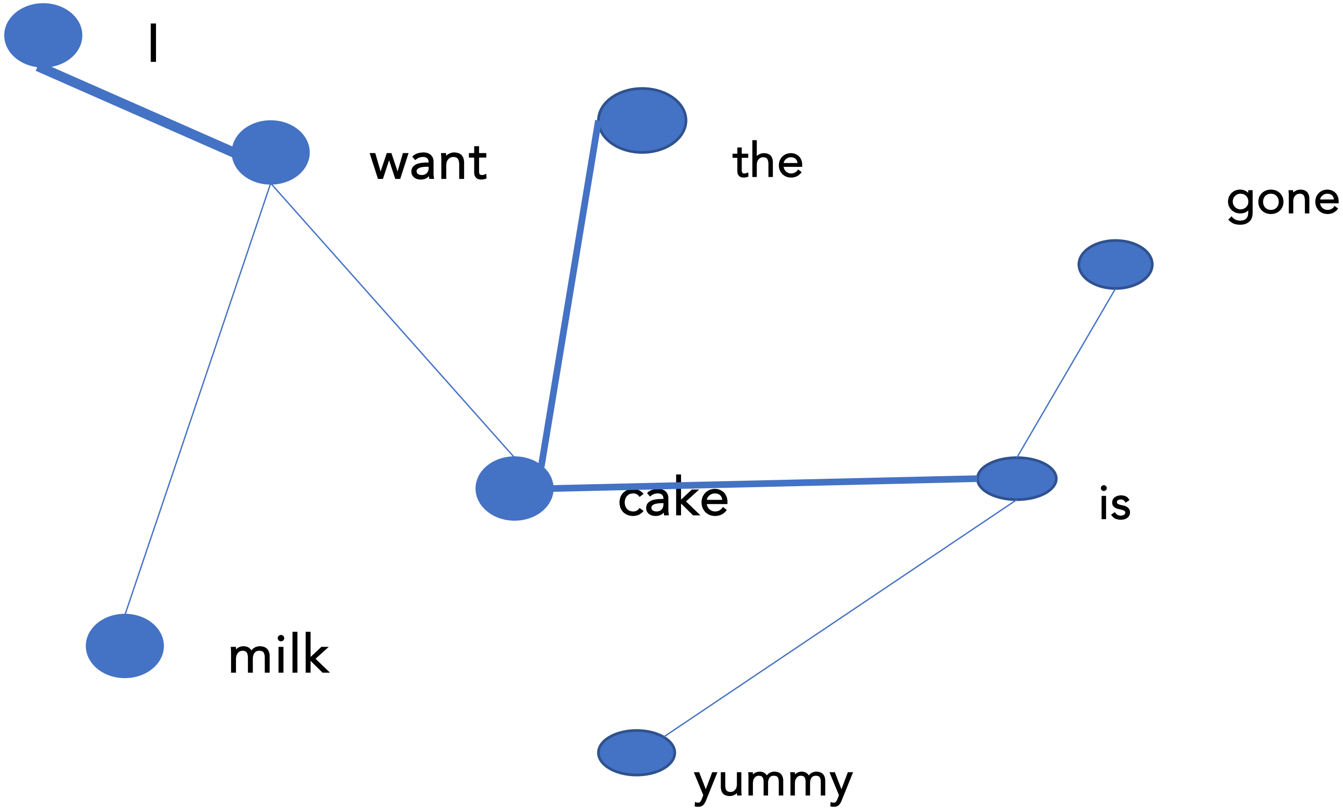


I want milk

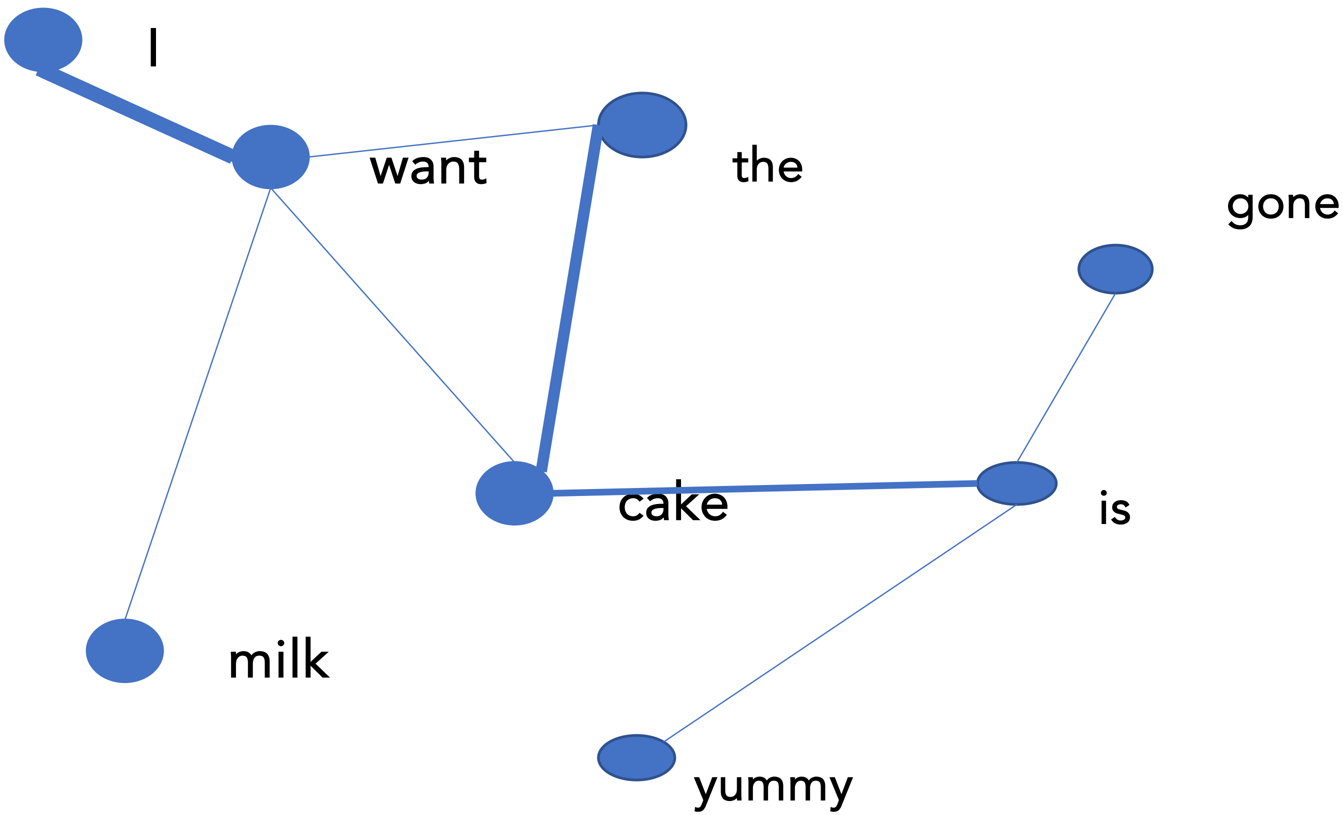
I want cake



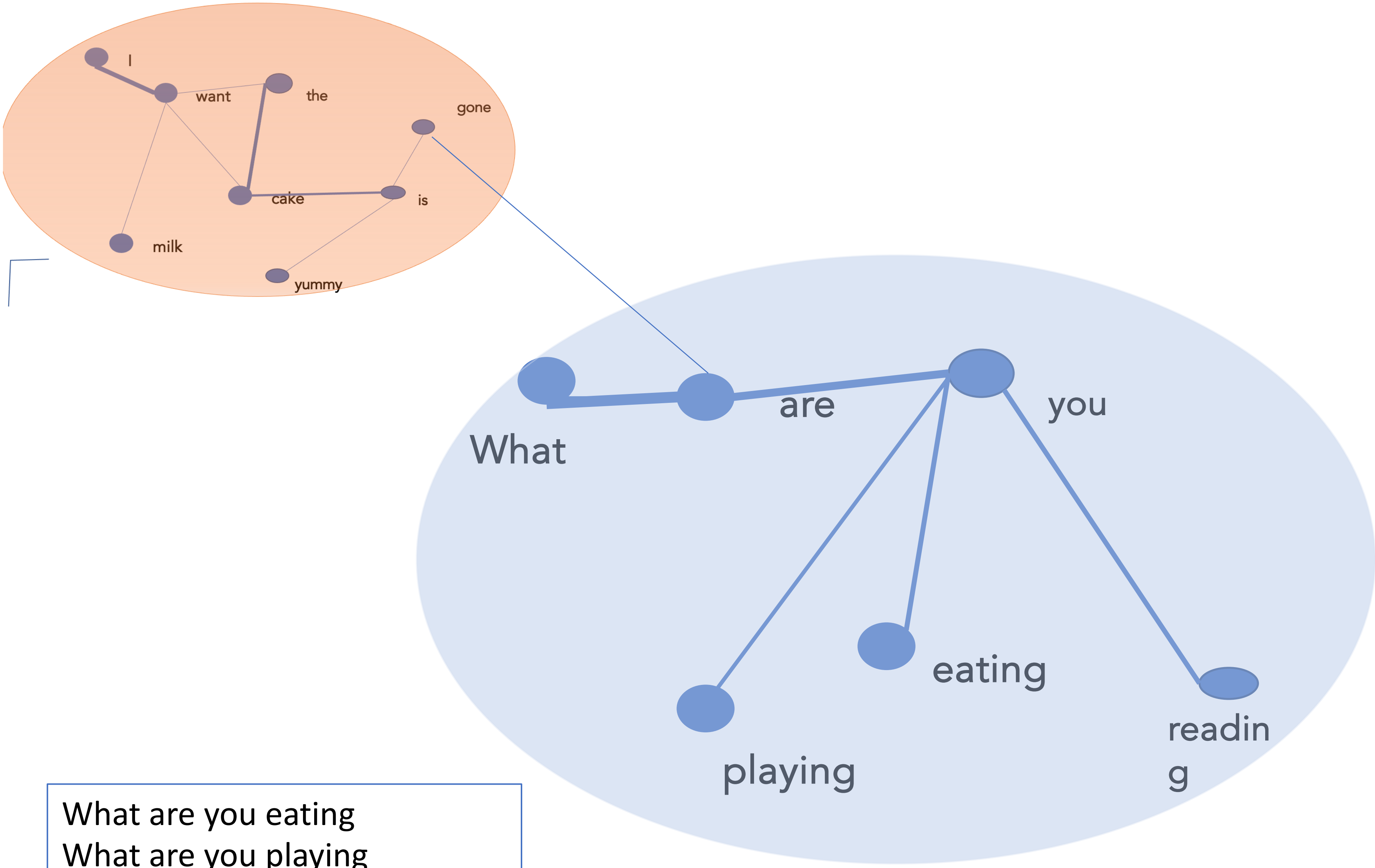
The cake is yummy.
The cake is gone.



The cake is yummy.
The cake is gone.



I want the cake



What are you eating
What are you playing
What are you reading
What are you x-ing

Data - Bilingual Corpora

Silvie

2;4 - 3;10

N_{CHI} = 65.473

N_{Input}= 140.387

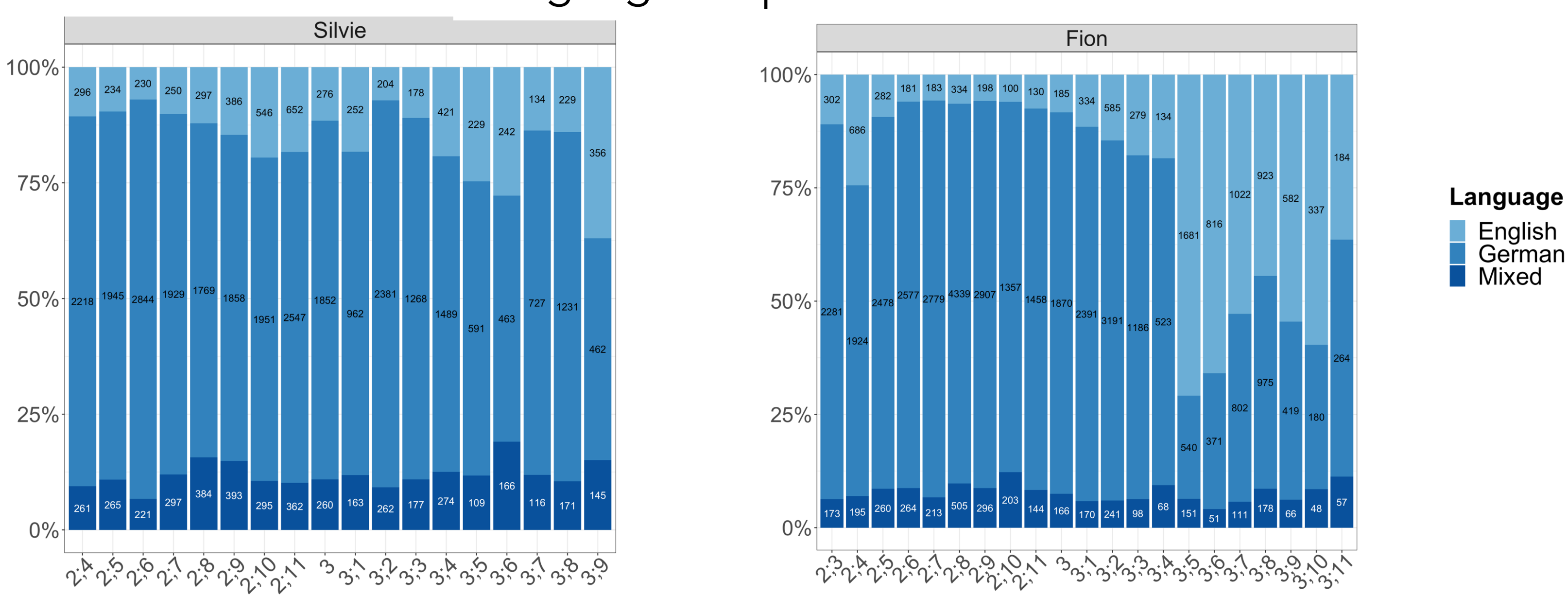
Fion

2;3 - 3;11

N_{CHI} = 47.928

N_{input} = 180.292

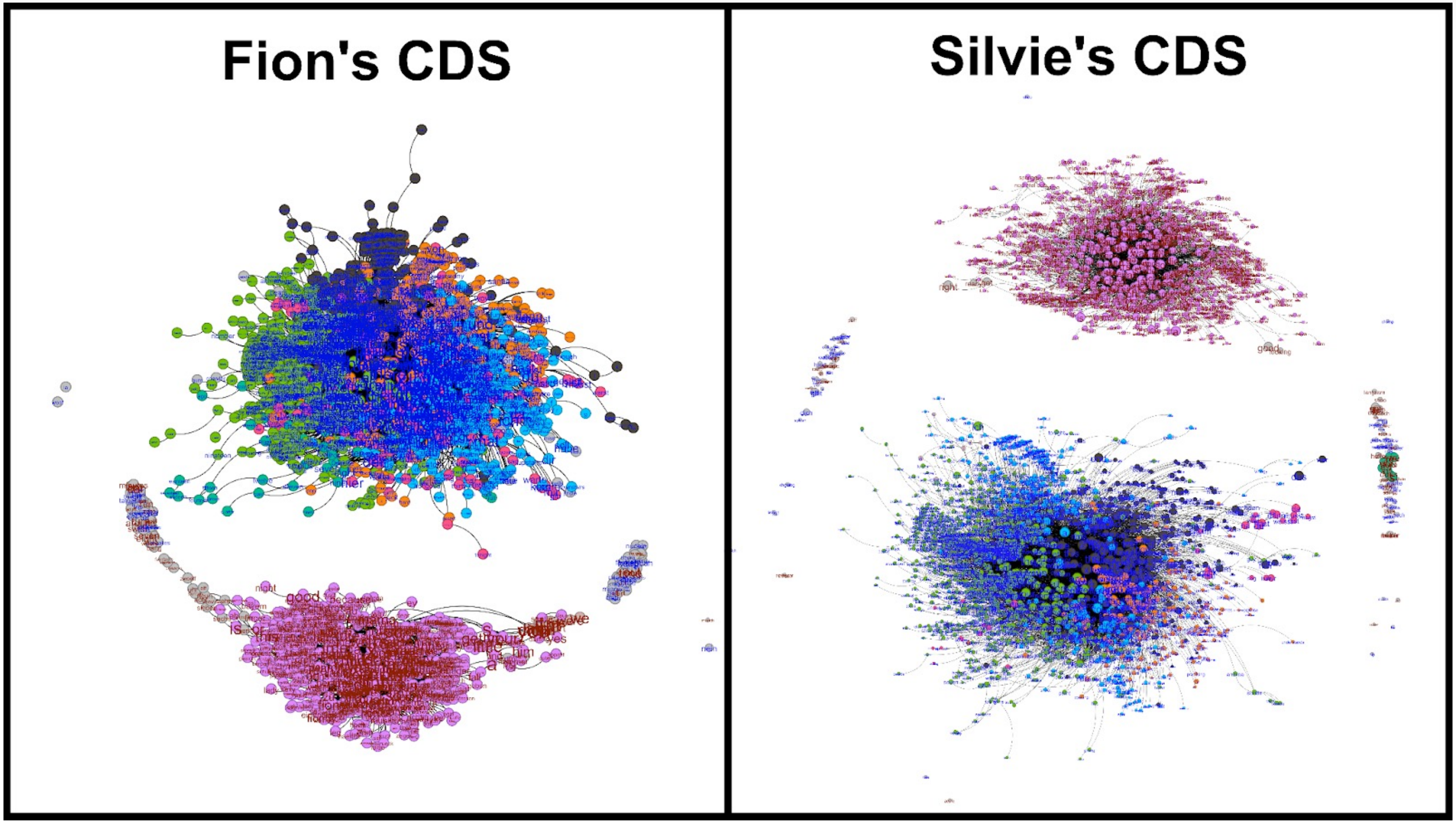
Language Proportions



Constructing a DNM

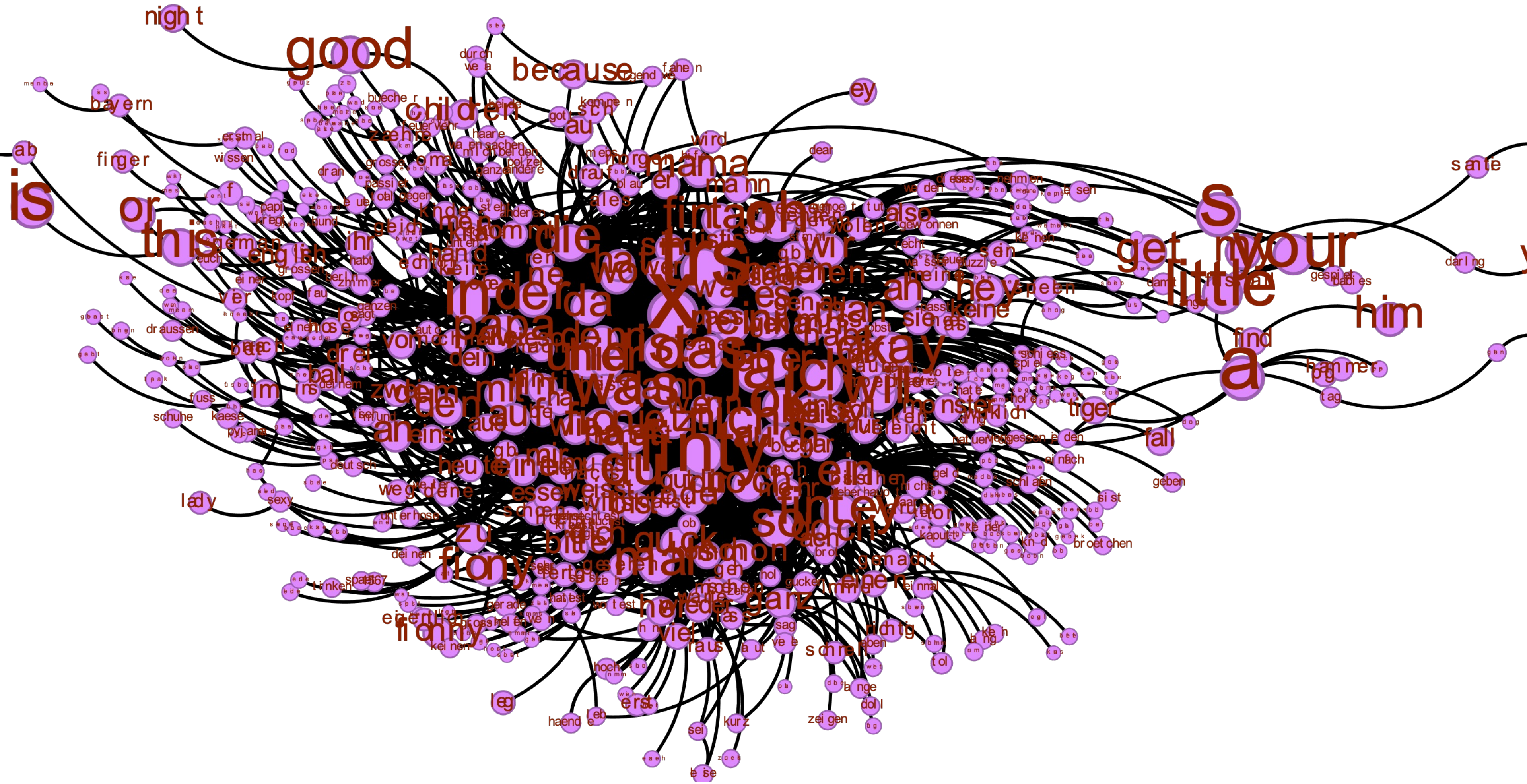
- ✓ Breaking down the child data into bigrams (as Ibbotson et al. 2019 did for CDS)
- ✓ Constructing a network from transitional probabilities between each two words (using igraph, Csardi & Nepusz 2006)
- ✓ Community detection using the Louvain method (Blondel et al. 2008) as implemented in Gephi (Bastian et al. 2009)
- ✓ Visualization using Gephi

Results CDS

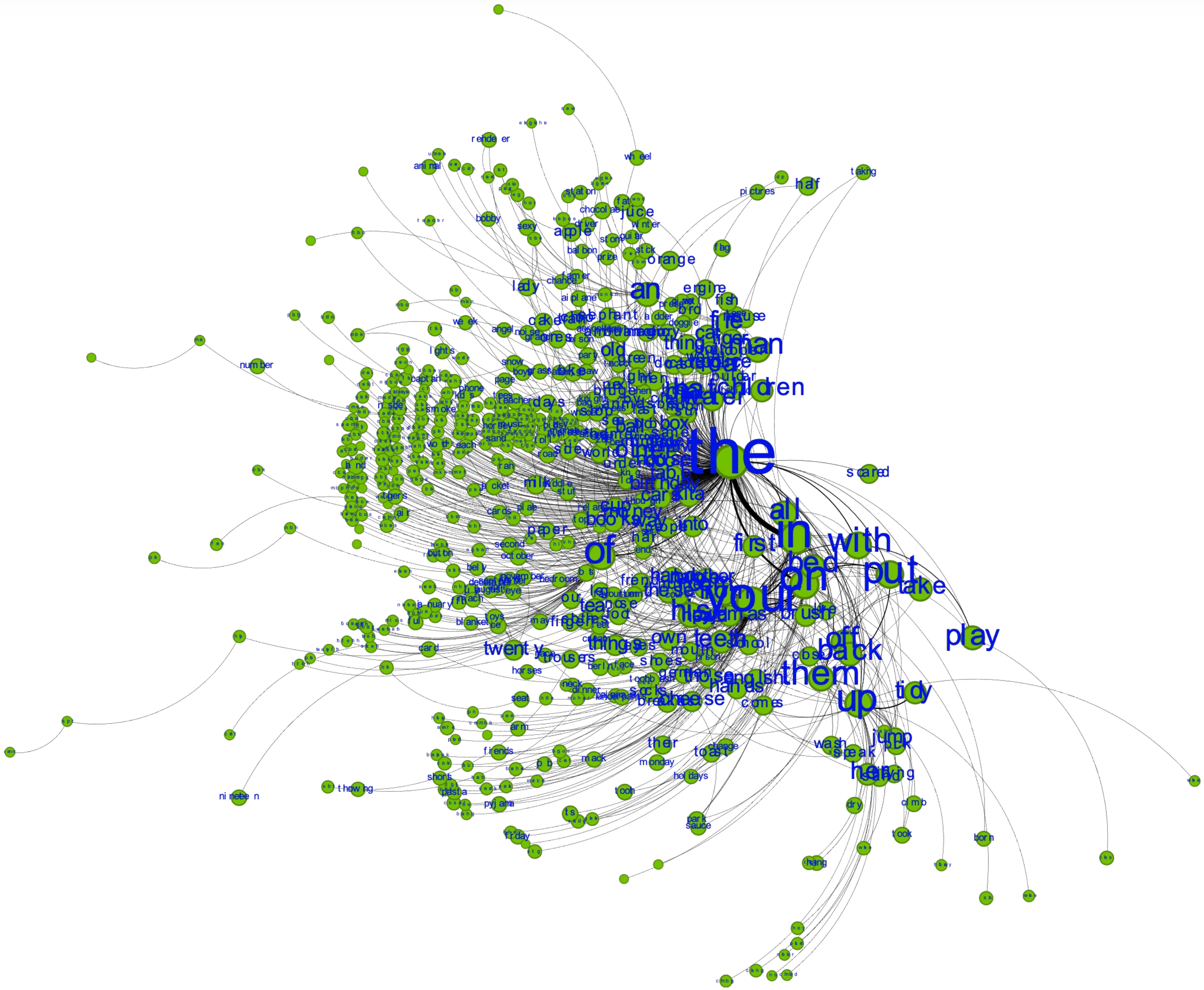


Results of a community detection algorithm run over the bigrams attested at least 5 times in the child-directed speech of Fion and Silvie. Red labels indicate German

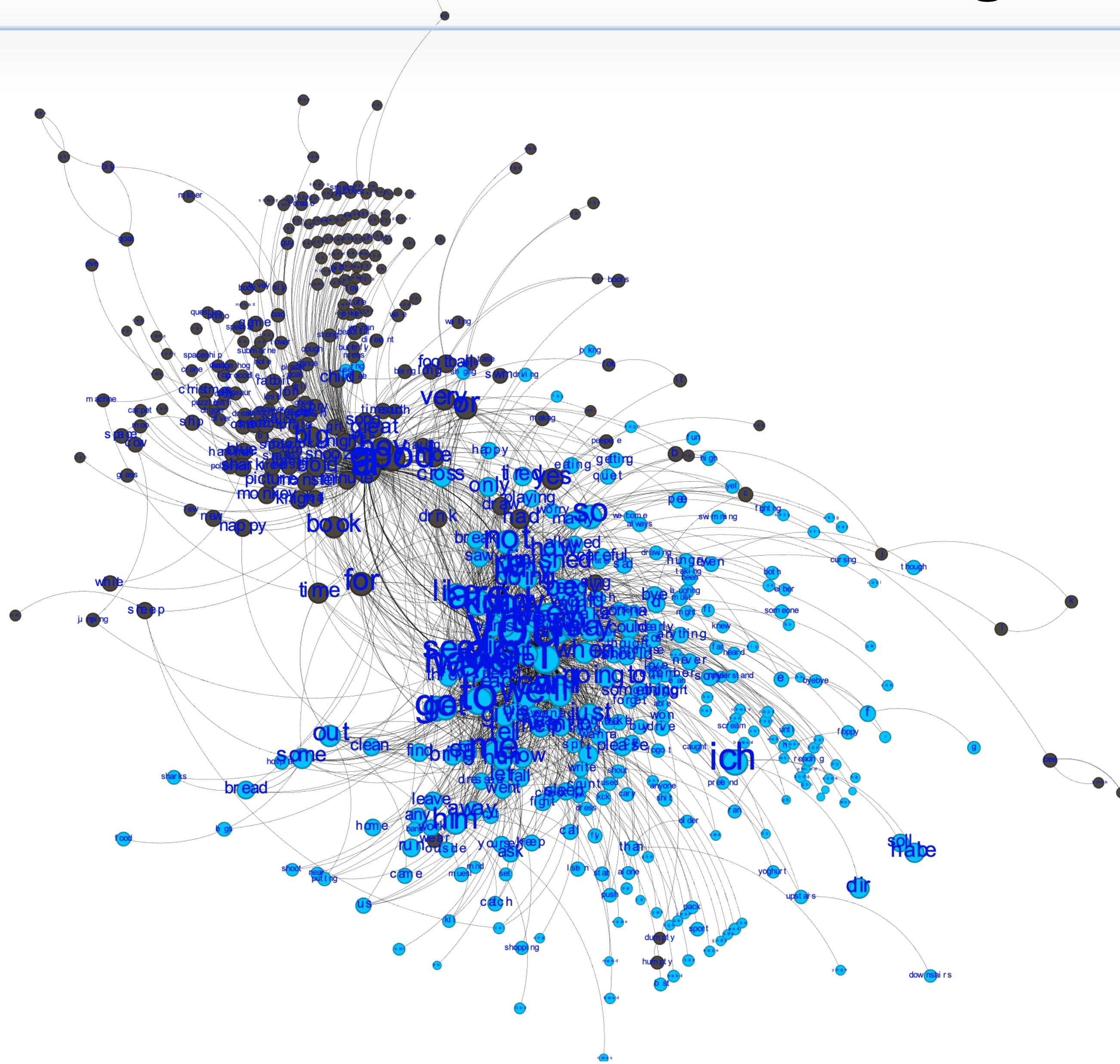
Fion CDS: Largest clusters



Fion CDS: Largest clusters

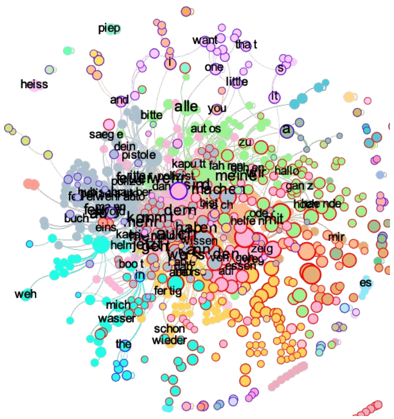


Fion CDS: Largest clusters

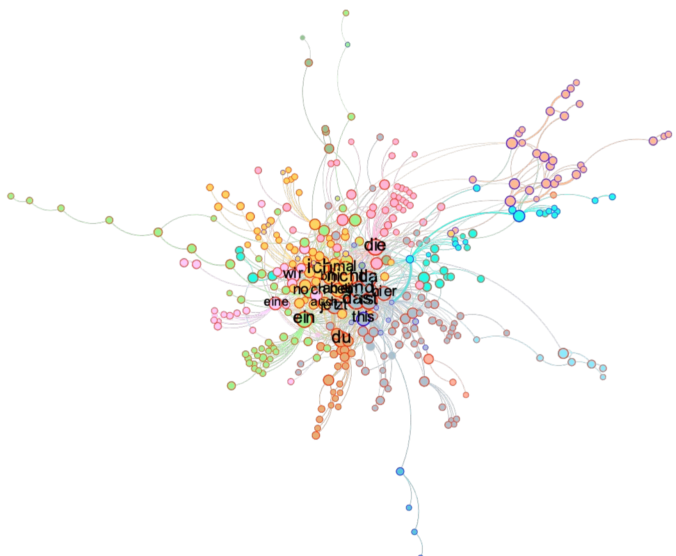


Results Child Speech: Fion

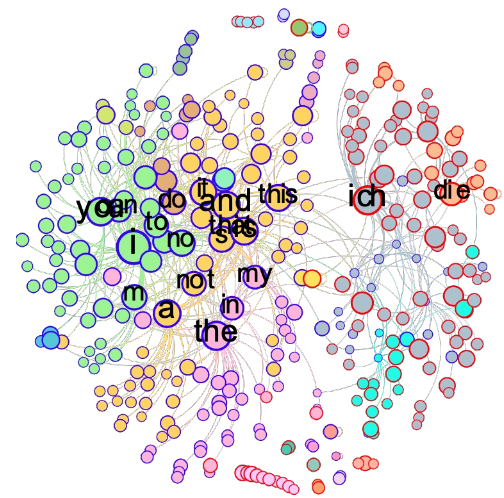
02;03-02;09



02;10-03;04

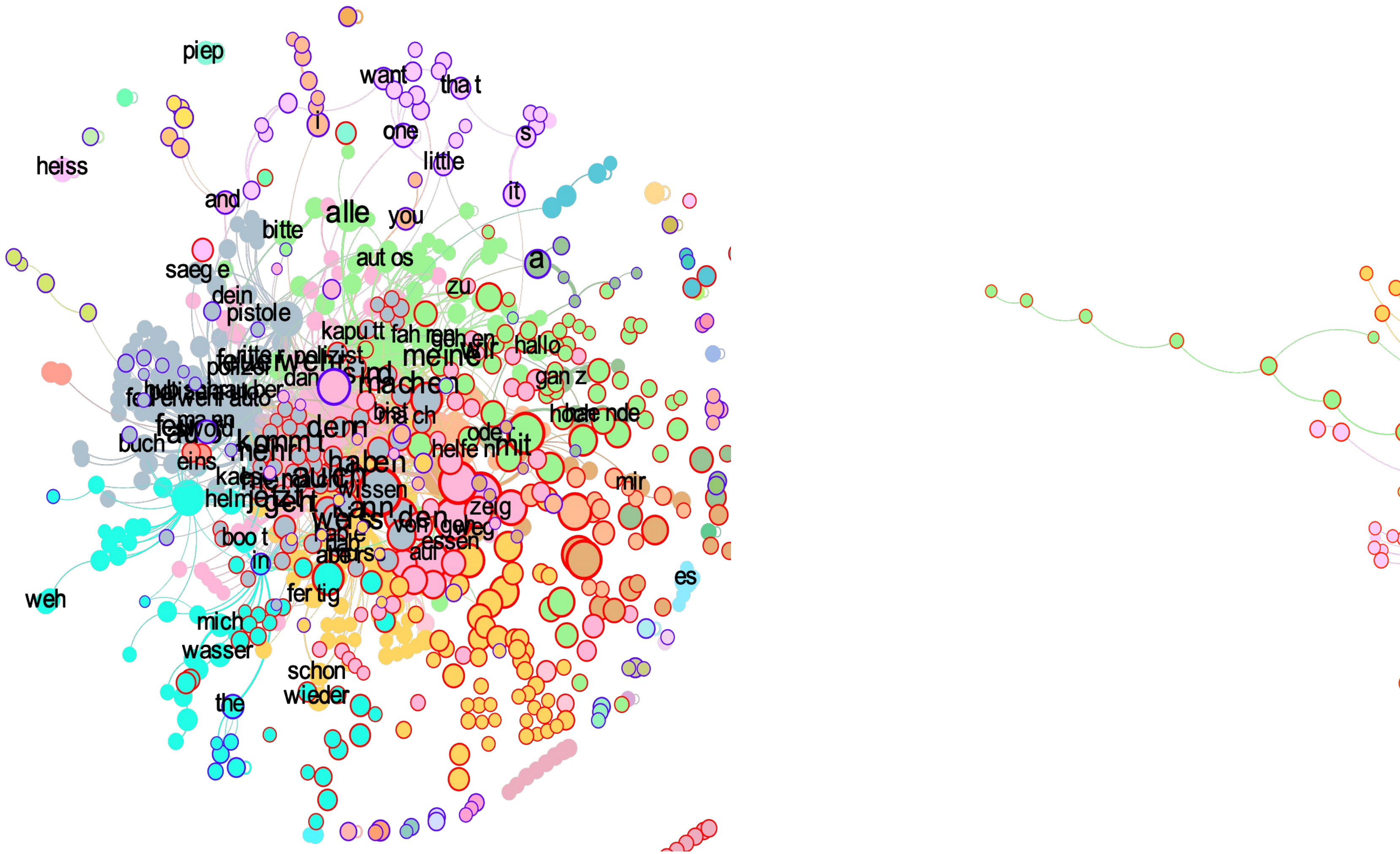


03;05-03;11



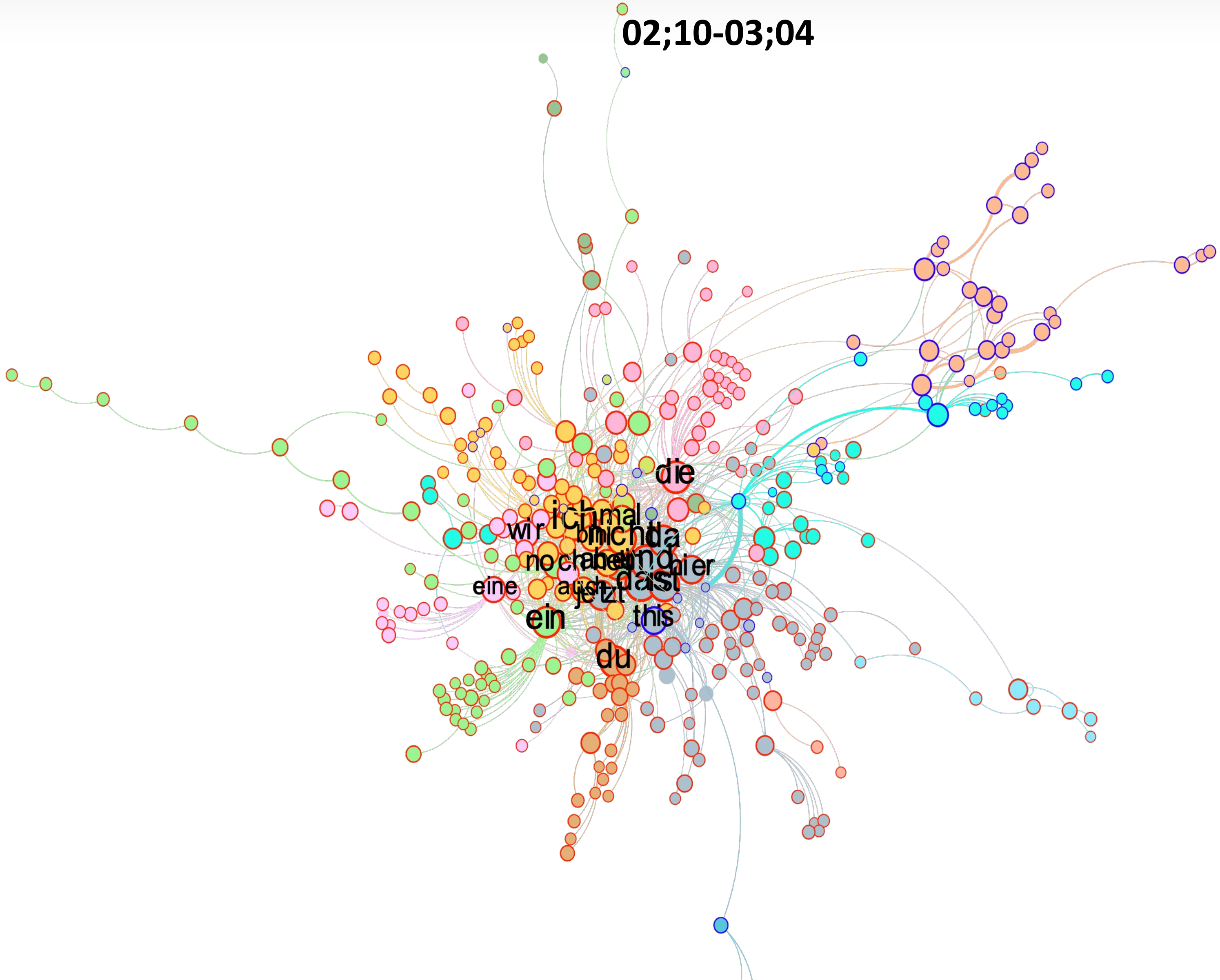
Results Child Speech: Fion

02;03-02;09



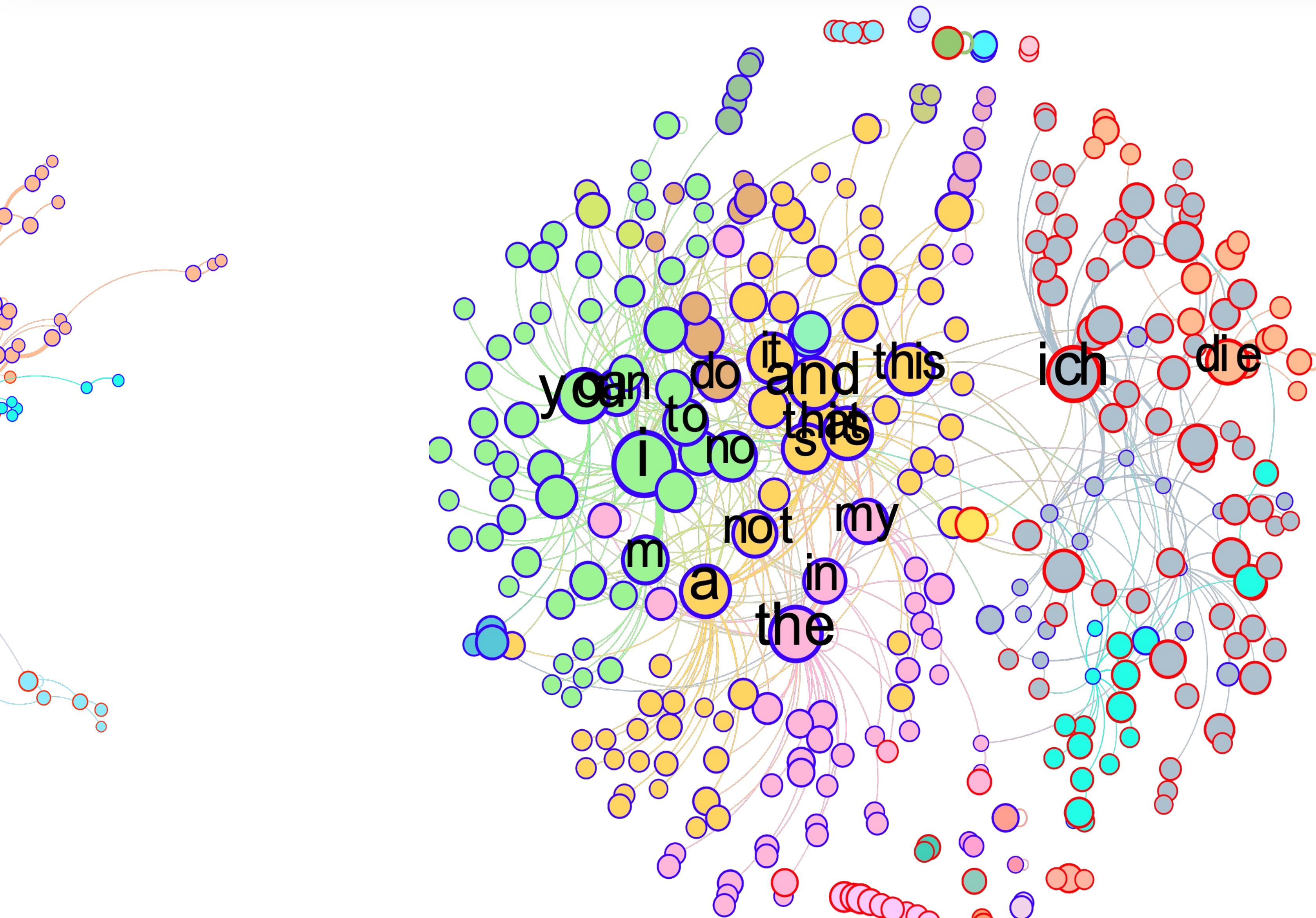
Results Child Speech: Fion

02;10-03;04



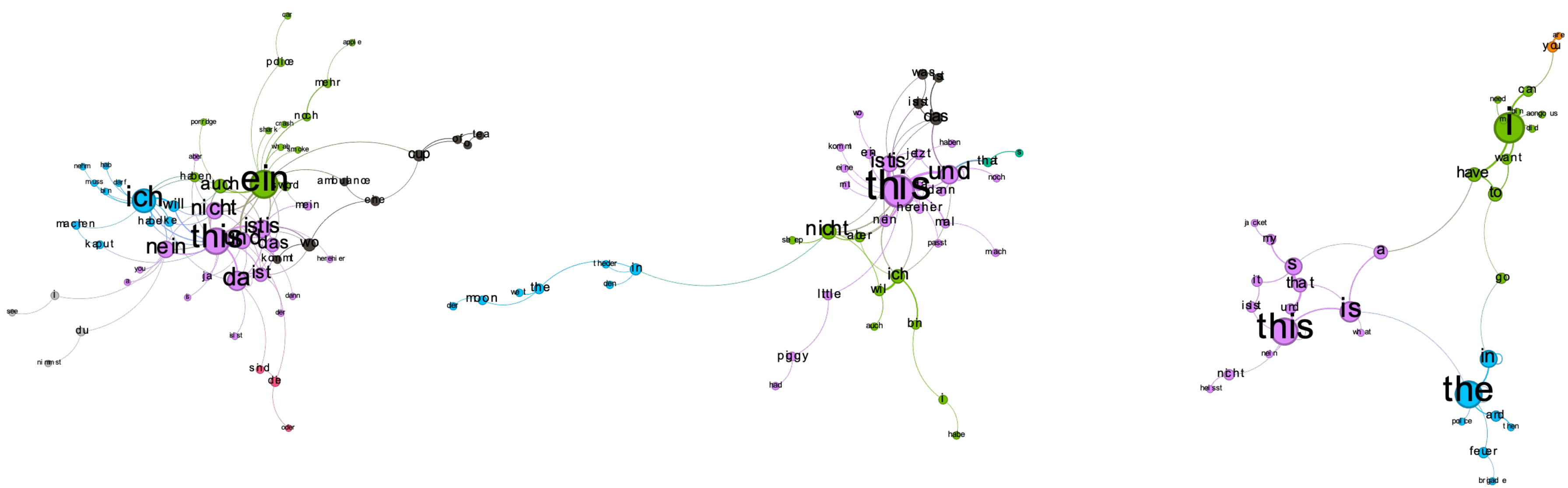
Results Child Speech: Fion

03;05-03;11



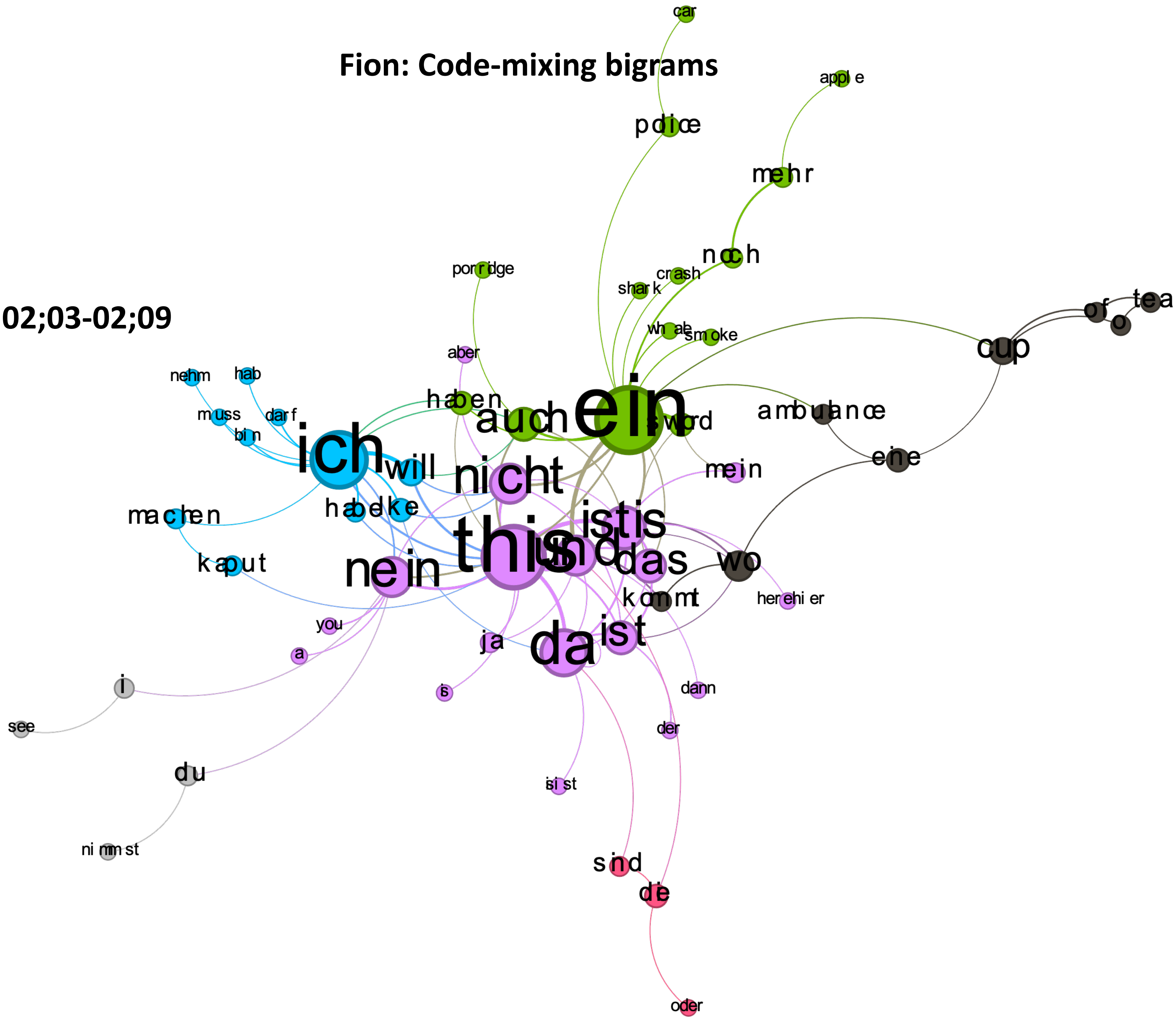
Fion: Code-mixing bigrams

02;03-02;09 | 02;10-03;04 | 03;05-03;11



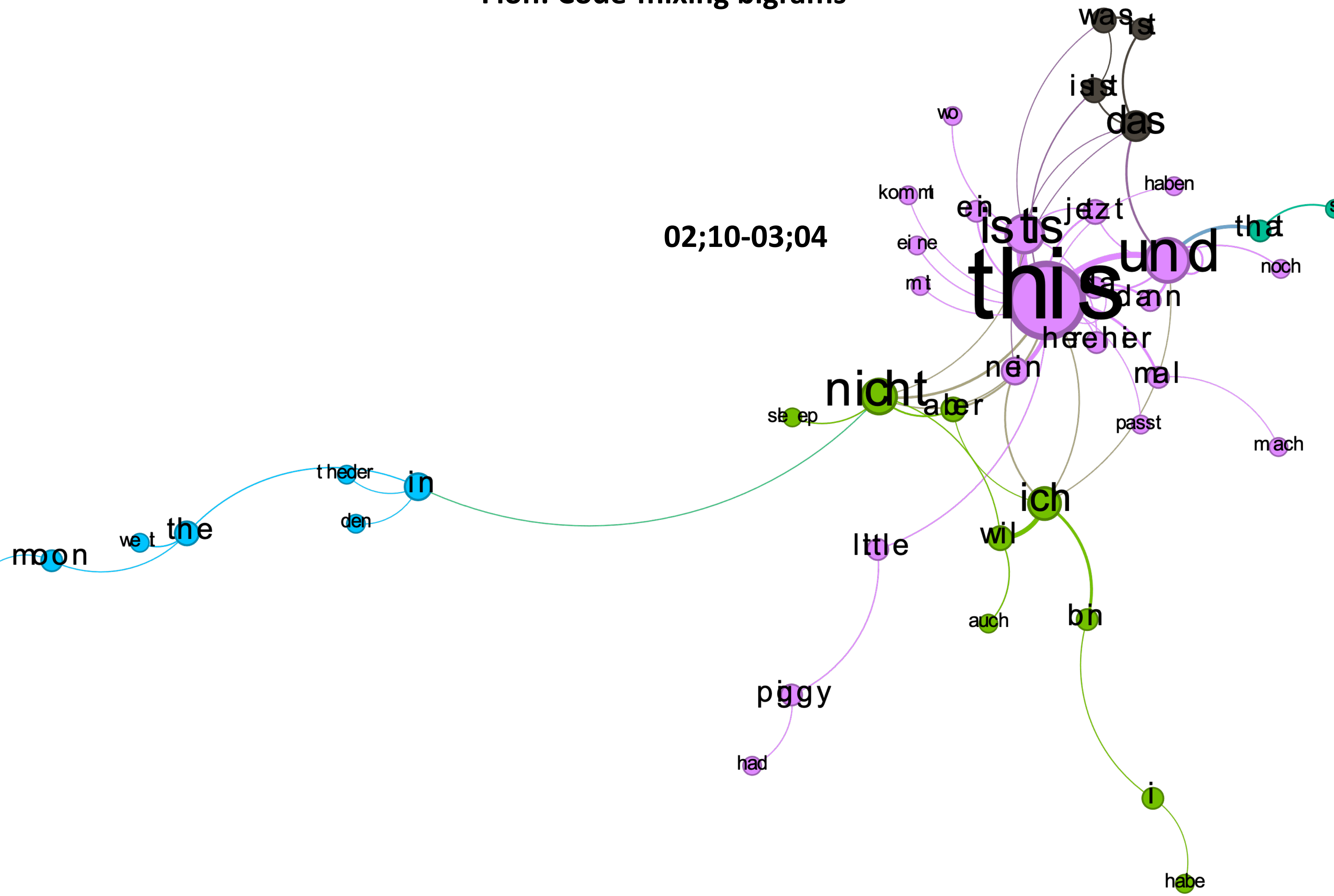
Fion: Code-mixing bigrams

02;03-02;09

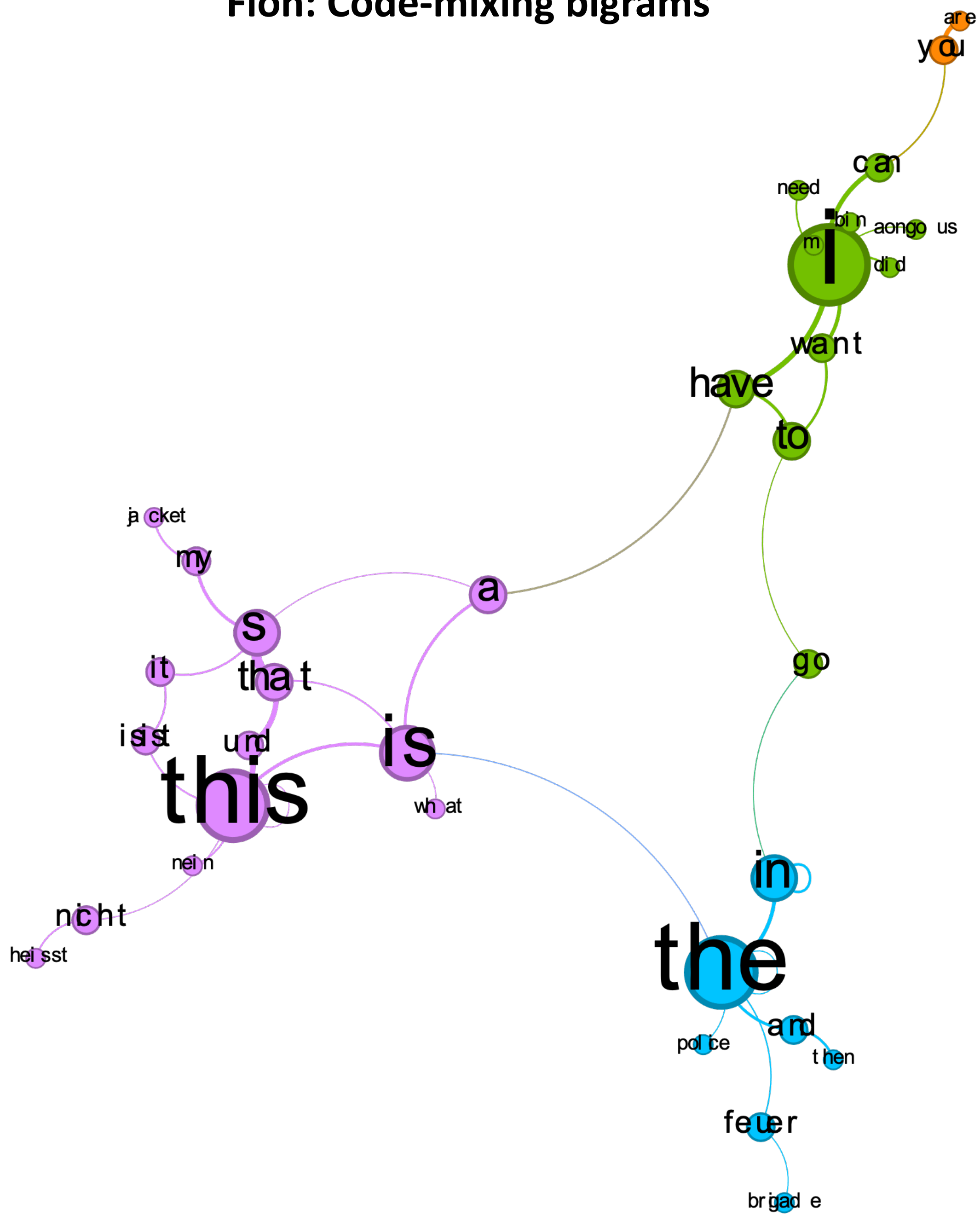


Fion: Code-mixing bigrams

02;10-03;04



Fion: Code-mixing bigrams



03;05-03;11

Conclusion

- Dynamic network algorithm from Ibbotson et al. (2019) can be fruitfully applied to child data
- "there are several tightly interconnected clusters with some nodes acting as bridges or hubs to other densely connected clusters" (Ibbotson et al. 2019)
- OPOL input forms clearly distinct "islands" → forced language separation
- Child data displays "hubs" reflecting patterns of use
- in the case of code-mixing data, it points to the "pivots" that are particularly important for language switches
- in the case of Fion's data, they also point to the re-organization of his "network" due to his shift from German to English
- The network "hubs" show how the code-mixing patterns change over time.

holistic networks capture the full linguistic repertoire


Спасибо
Gracias
شكرا
Obrigado
Спасибо
Dank U
Grazie
Ευχαριστώ
Danke
Ευχαριστώ
Merci
Thank You
Ngiyabonga
Dank U
Diolch
Obrigado
Ngiyabonga
Thank You
Merci
Tack
Ευχαριστώ
Dzięki
Danke
Grazie
תודה
Dank U
Terima Kasih
Diolch
Grazie
Dzięki
Merci
Diolch
Gracias
Tack
Tack
תודה

Literature

- Behrens, Heike. 2006. The input–output relationship in first language acquisition. *Language and Cognitive Processes* 21(1–3). 2–24.
- Cameron-Faulkner, Thea, Elena Lieven & Michael Tomasello. 2003. A construction based analysis of child directed speech. *Cognitive Science* 27(6). 843–873.
- Dąbrowska, Ewa & Elena Lieven. 2005. Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16(3). 437–474.
- Ibbotson, Paul, Vsevolod Salnikov & Richard Walker. 2019. A dynamic network analysis of emergent grammar. *First Language* 39(6). 652–680.
- Koch, Nikolas. 2019. *Schemata im Erstspracherwerb: eine Traceback-Studie für das Deutsche*. (Linguistik, Impulse & Tendenzen 80). Berlin, Boston: De Gruyter.
- Koch, Nikolas, Stefan Hartmann & Antje Endesfelder Quick. 2020a. The traceback method and the early constructicon: theoretical and methodological considerations. *Corpus Linguistics and Linguistic Theory* 0(0). doi:10.1515/cllt-2020-0045.
- Koch, Nikolas, Stefan Hartmann & Antje Endesfelder Quick. 2020b. The traceback method and the early constructicon: theoretical and methodological considerations. *Corpus Linguistics and Linguistic Theory* 0(0). 000010151520200045.
- Kol, Sheli, Bracha Nir & Shuly Wintner. 2014. Computational evaluation of the Traceback Method. *Journal of Child Language* 41(1). 176–199.
- Lieven, Elena, Heike Behrens, Jennifer Speares & Michael Tomasello. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30. 333–370.
- Lieven, Elena, Dorothé Salomo & Michael Tomasello. 2009. Two-year-old Children's Production of Multiword Utterances: A Usage-Based Analysis. *Cognitive Linguistics* 20. 481–507.
- Quick, Antje Endesfelder, Ad Backus & Elena Lieven. 2021. Entrenchment effects in code-mixing: individual differences in German-English bilingual children. *Cognitive Linguistics* 0(0). doi:10.1515/cog-2020-0036.
- Quick, Antje Endesfelder & Stefan Hartmann. 2021. The Building Blocks of Child Bilingual Code-Mixing: A Cross-Corpus Traceback Approach. *Frontiers in Psychology* 12. 2203.
- Quick, Antje Endesfelder, Stefan Hartmann, Ad Backus & Elena Lieven. 2021. Entrenchment and productivity: The role of input in the code-mixing of a German-English bilingual child. *Applied Linguistics Review* 12(2). 225–247.
- Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge and London: Harvard University Press.
- Vogt, Paul & Elena Lieven. 2010. Verifying Theories of Language Acquisition Using Computer Models of Language Evolution. *Adaptive Behavior* 18(1). 21–35.