- Parallel Workshop 2-

The DIPEx database and NLP approaches to analysis
(English)

Workshop Leaders: Gerold Schneider, Giovanni Spitale, Tilia Ellendorff

gschneid@ifi.uzh.ch
giovanni.spitale@ibme.uzh.ch
tilia.ellendorff@cl.uzh.ch

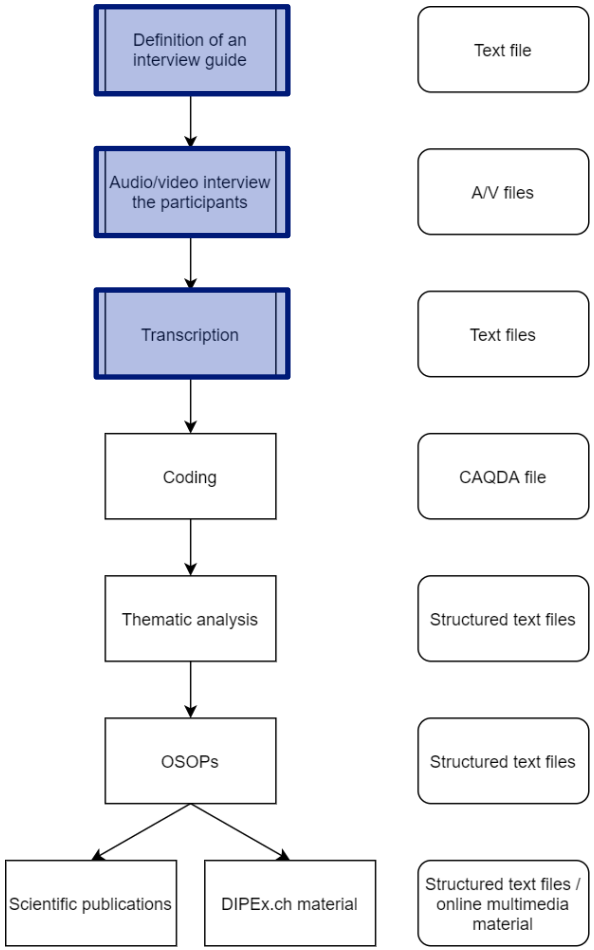
Content WS 2

1. Introduction Input
2. Leading questions for discussion
3. Output of discussion

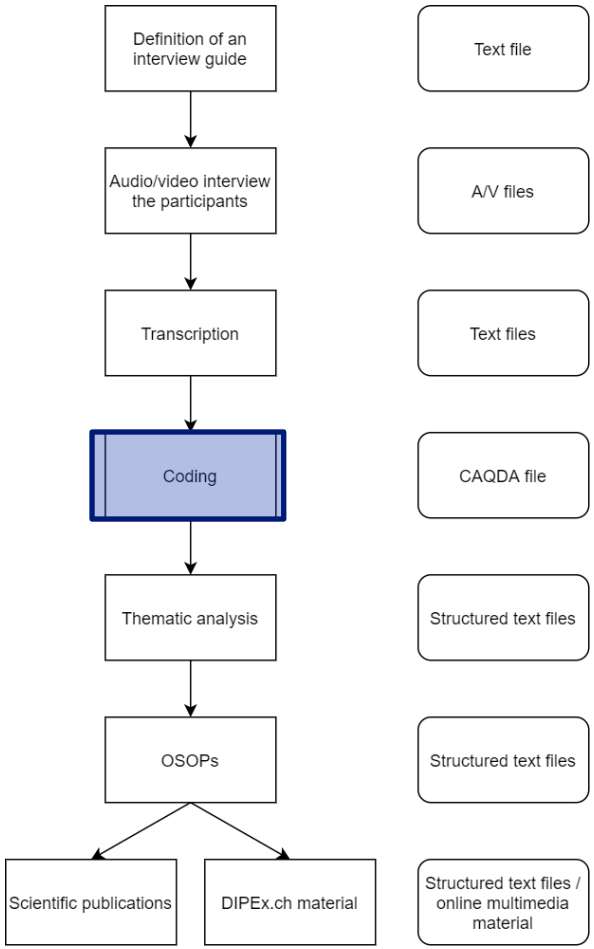
1. Input
1.1 The Dipex Database

Interview guide
Semi-structured document listing the questions to ask and the prompts to give to the interviewee. Starts with an open section then follows specific topics of interest.

Interview files
The interview is audio and/or video recorded – according to the preferences of the itwee – and then transcribed as text.

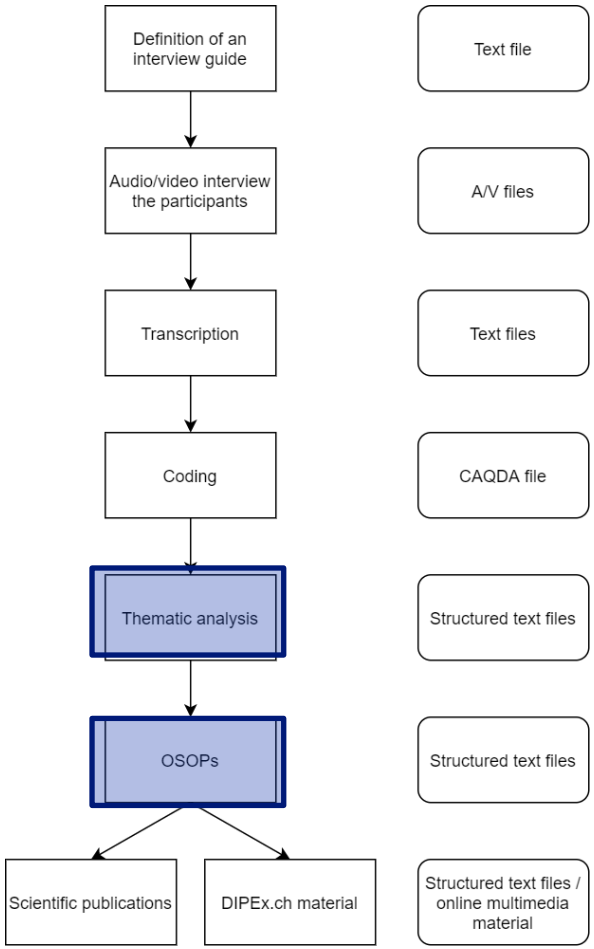


1. Input
1.1 The Dipex Database



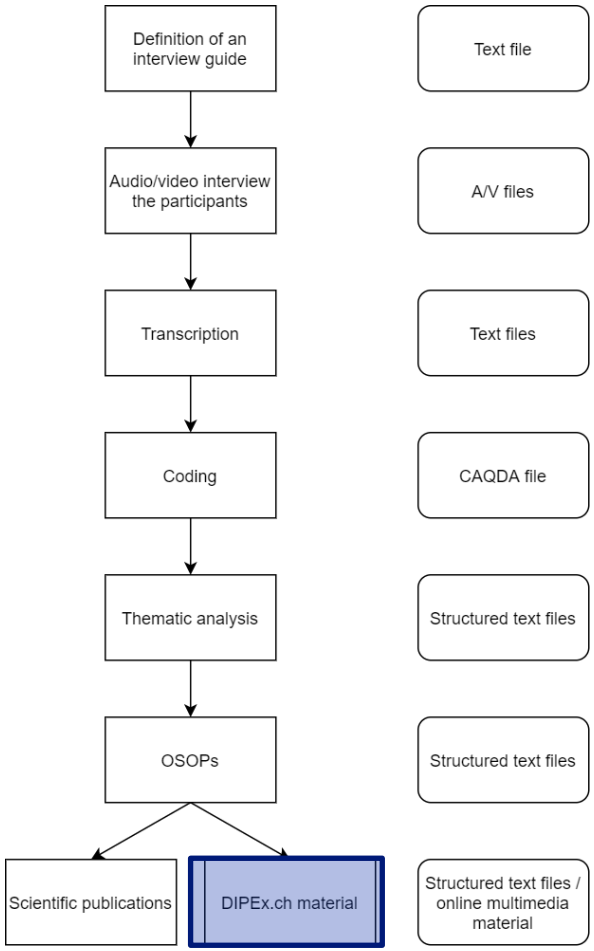
Coding
The interviews are loaded in a software for computer assisted qualitative data analysis. We define a specific coding tree and manually code the text (= assign one or more labels to a meaningful passage)

1. Input
1.1 The Dipex Database



Thematic analysis and OSOPs
We select specific topics that allow us to tell the 'collective history' of a given experience weaving individual voices together. We attribute codes to topics, retrieve the quotes, and put the story together.

1. Input
1.1 The Dipex Database

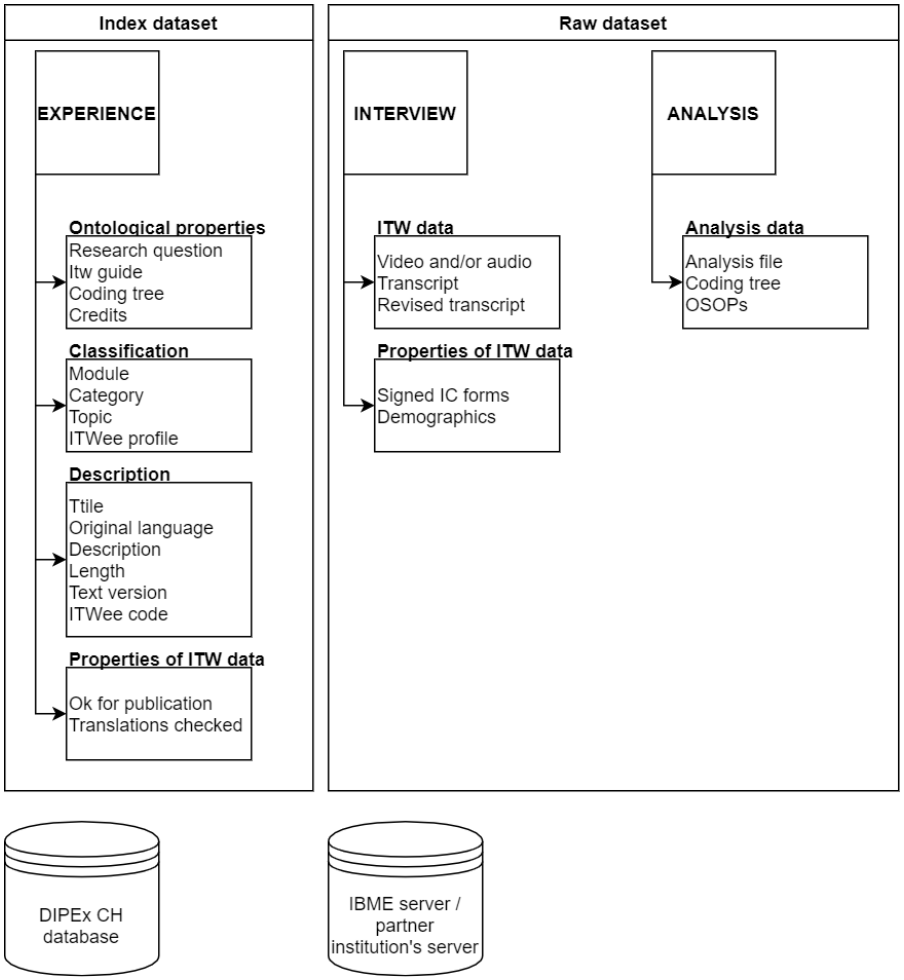


Output
Based on the OSOPs, we prepare selected material to be put online and serve as an important resource for patients, relatives, caregivers, healthcare professionals, and students.

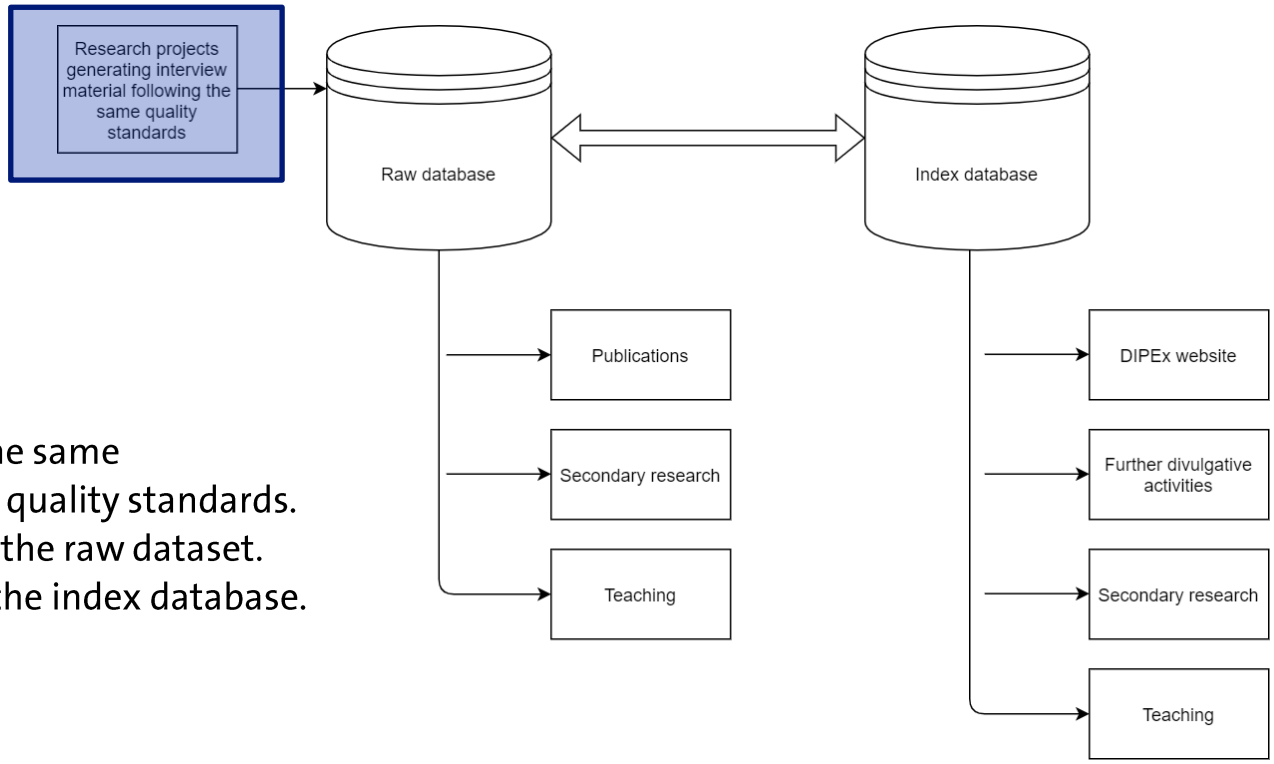
1. Input
1.1 The Dipex Database

Raw dataset
The DNA of our research data; located on IBME's servers

Index dataset
The index and mRNA of our research data (extended metadata); located on UZH's MariaDB instance

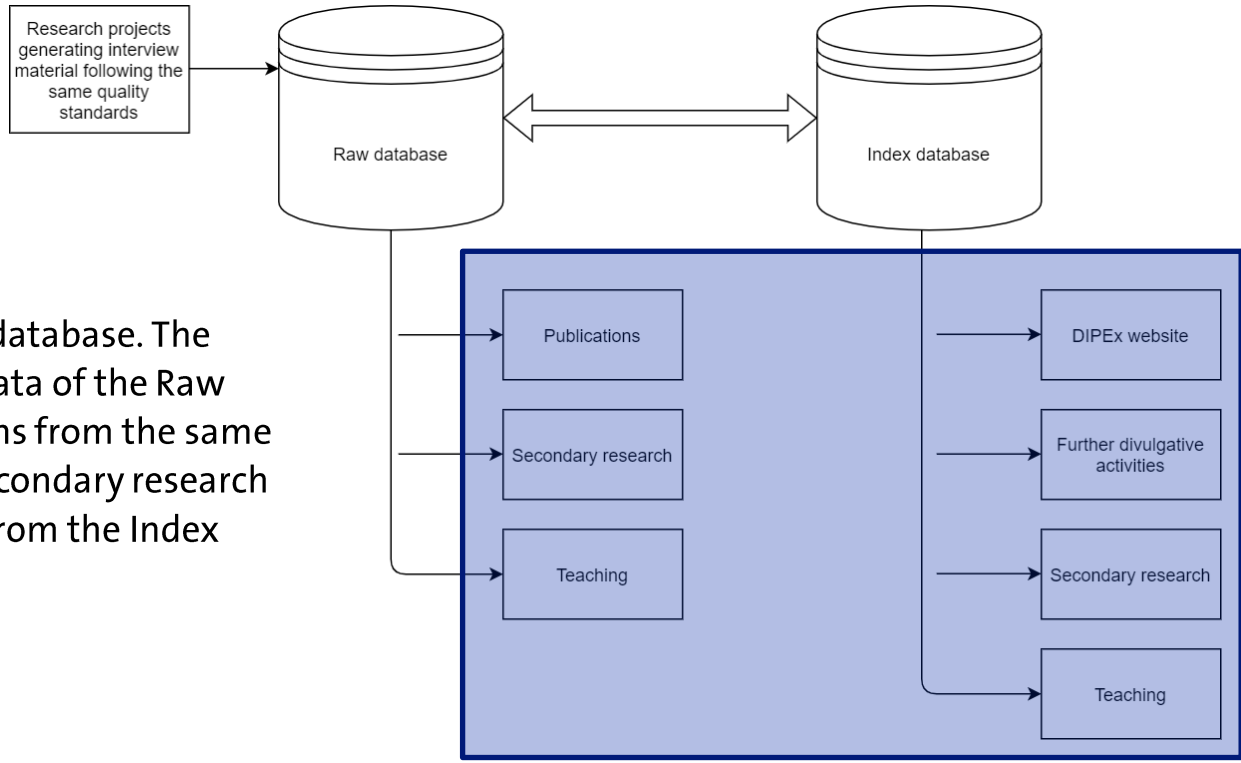


1. Input
1.1 The Dipex Database



Data in:
Interview data generated with the same methodology and with the same quality standards.
Raw interview data are stored in the raw dataset.
Their content is then indexed in the index database.

1. Input
1.1 The Dipex Database



Data out:
Data can be explored via the Index database. The Index database contains the metadata of the Raw database, but also organized sections from the same data. Depending on the purpose, secondary research or teaching activities can use data from the Index dataset or from the Raw dataset.

1. Input

1.1 The Dipex Database

Filter experiences by topic

This creates a slice of the dataframe containing only the experiences belonging to a specific topic

```
[23]: experience_df_subset[experience_df_subset['EXP_topic'] == 'Delusion/hallucination/ dreams/anxiety.' ]
```

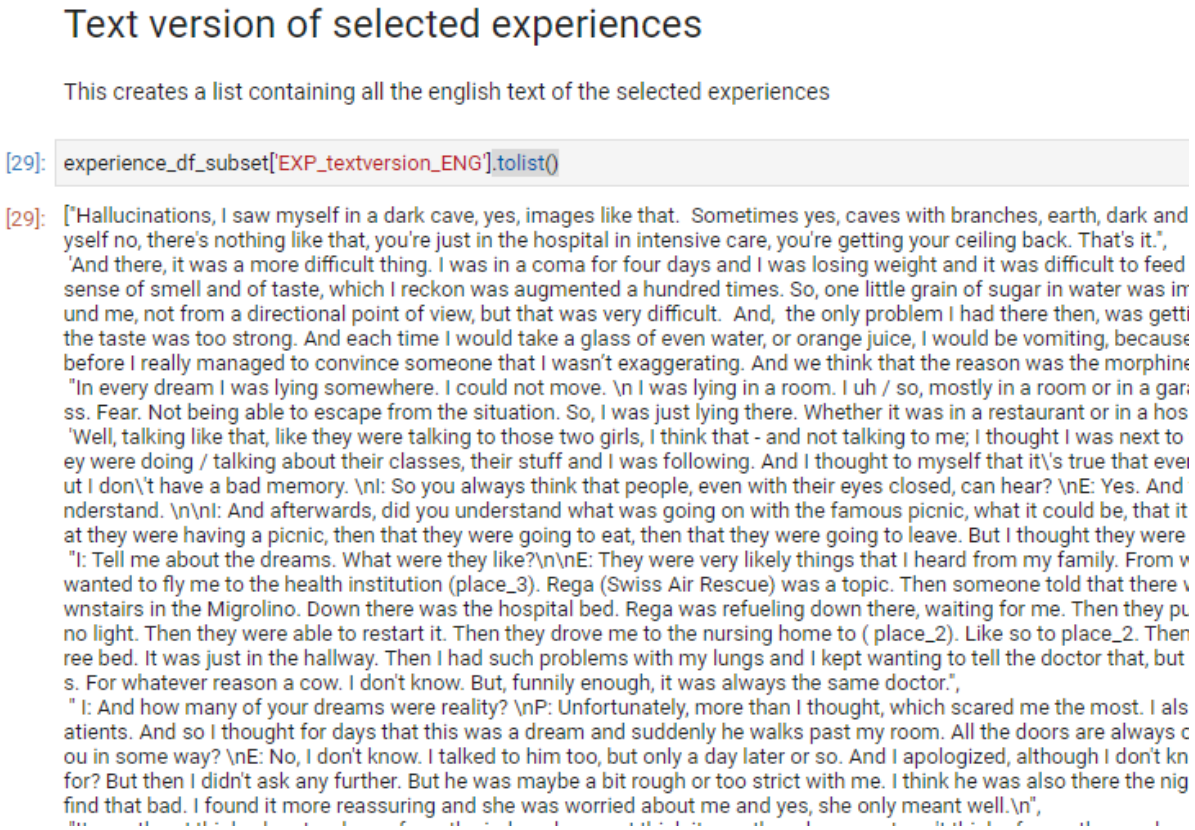
[23]:

	ID_Experience	Belongs_to_module	ITW_code	ITWee_code	ITW_original_language	EXP_start_time	EXP_end_time	EXP_duration
21	31	CMI	CMI01_22102019_FR	CMI01	FRE	0 days 00:32:02	0 days 00:32:45	0 days 00:00:43
22	32	CMI	CMI03_05122019_EN	CMI03	ENG	0 days 00:05:35	0 days 00:07:20	0 days 00:01:45
23	33	CMI	CMI08_16122019_DE	CMI08	DEU	0 days 00:06:06	0 days 00:07:06	0 days 00:01:00
24	34	CMI	CMI05_05122019_FR	CMI05	FRE	0 days 00:27:48	0 days 00:29:16	0 days 00:01:28
25	35	CMI	CMI10_15012020_DE	CMI10	DEU	0 days 00:06:43	0 days 00:08:20	0 days 00:01:37
26	36	CMI	CMI20_29022020_DE	CMI20	DEU	NaT	NaT	NaT
27	37	CMI	CMI23_03072020_FR	CMI23	FRE	0 days 00:20:30	0 days 00:21:18	0 days 00:00:48

Getting all the experiences belonging to a topic
...in 1 line of code

1. Input
1.1 The Dipex Database

Getting all the text (in English) of the selected experiences
...in 1 line of code



Text version of selected experiences

This creates a list containing all the english text of the selected experiences

```
[29]: experience_df_subset['EXP_textversion_ENG'].tolist()
```

[29]: ["Hallucinations, I saw myself in a dark cave, yes, images like that. Sometimes yes, caves with branches, earth, dark and
yself no, there's nothing like that, you're just in the hospital in intensive care, you're getting your ceiling back. That's it.",
'And there, it was a more difficult thing. I was in a coma for four days and I was losing weight and it was difficult to feed
sense of smell and of taste, which I reckon was augmented a hundred times. So, one little grain of sugar in water was im
und me, not from a directional point of view, but that was very difficult. And, the only problem I had there then, was getti
the taste was too strong. And each time I would take a glass of even water, or orange juice, I would be vomiting, because
before I really managed to convince someone that I wasn't exaggerating. And we think that the reason was the morphine
"In every dream I was lying somewhere. I could not move. \n I was lying in a room. I uh / so, mostly in a room or in a gar
ss. Fear. Not being able to escape from the situation. So, I was just lying there. Whether it was in a restaurant or in a hos
'Well, talking like that, like they were talking to those two girls, I think that - and not talking to me; I thought I was next to
ey were doing / talking about their classes, their stuff and I was following. And I thought to myself that it\'s true that ever
ut I don\'t have a bad memory. \nI: So you always think that people, even with their eyes closed, can hear? \nE: Yes. And
nderstand. \n\nI: And afterwards, did you understand what was going on with the famous picnic, what it could be, that it
at they were having a picnic, then that they were going to eat, then that they were going to leave. But I thought they were
"I: Tell me about the dreams. What were they like?\n\nE: They were very likely things that I heard from my family. From w
wanted to fly me to the health institution (place_3). Rega (Swiss Air Rescue) was a topic. Then someone told that there w
wnstairs in the Migrolino. Down there was the hospital bed. Rega was refueling down there, waiting for me. Then they pu
no light. Then they were able to restart it. Then they drove me to the nursing home to (place_2). Like so to place_2. Then
ree bed. It was just in the hallway. Then I had such problems with my lungs and I kept wanting to tell the doctor that, but
s. For whatever reason a cow. I don't know. But, funnily enough, it was always the same doctor.",
" I: And how many of your dreams were reality? \nP: Unfortunately, more than I thought, which scared me the most. I als
atients. And so I thought for days that this was a dream and suddenly he walks past my room. All the doors are always c
ou in some way? \nE: No, I don't know. I talked to him too, but only a day later or so. And I apologized, although I don't kn
for? But then I didn't ask any further. But he was maybe a bit rough or too strict with me. I think he was also there the nig
find that bad. I found it more reassuring and she was worried about me and yes, she only meant well.\n",

1.2 NLP Approaches to Analysis: Methods

How can we explore & analyse DipEx texts automatically?

Raw texts or annotated texts? → Supervised vs. Unsupervised

There are many useful methods. We just present three of them.

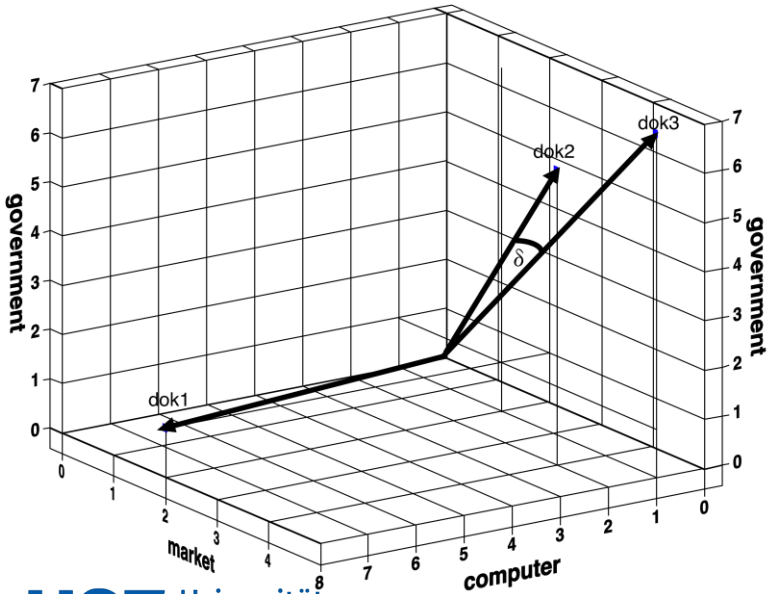
1. Document Classification (supervised)
2. Distributional Semantics (unsupervised)
3. Conceptual Maps (unsupervised or semi-supervised)

These approaches share that they need numerical text representations, typically vector spaces

1.2 NLP Approaches to Analysis: Vector Spaces

Document Classification uses Document-Term Matrices

Frequency	*market*	*computer*	*government*
dok1	2	8	1
dok2	4	2	6
dok3	5	1	7



- Context Windows e.g.[-10 w_0 +10] give us Term-Term Matrices

	dog	hyena	cat
runs	1	1	4
barks	5	2	0

TABLE 1 Distributional vectors representing the words *dog*, *hyena* and *cat*.

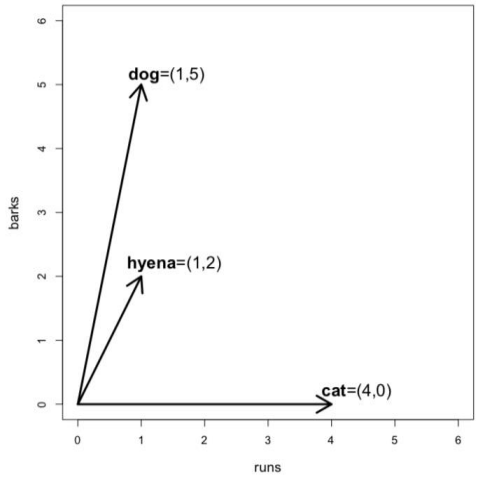


FIGURE 1 Geometric representation of the vectors in Table 1.

Source: Baroni et al. 2014:13

1.2 NLP Appraoches to Analysis: Teaser 2: Distributional Semantics with DIPEx MS

```
> closest_to(training_ms20,"leben")
          word similarity to "leben"
1         leben             1.0000000
2  gelassenheit             0.5548906
3   beziehungen             0.5310467
4      ausmacht             0.5218526
5      negative             0.5194965
6     krankheit             0.5168033
7        dingen             0.5158901
8    lebenswert             0.5157932
9        lebens             0.5121614
10    positiven             0.5043837
```

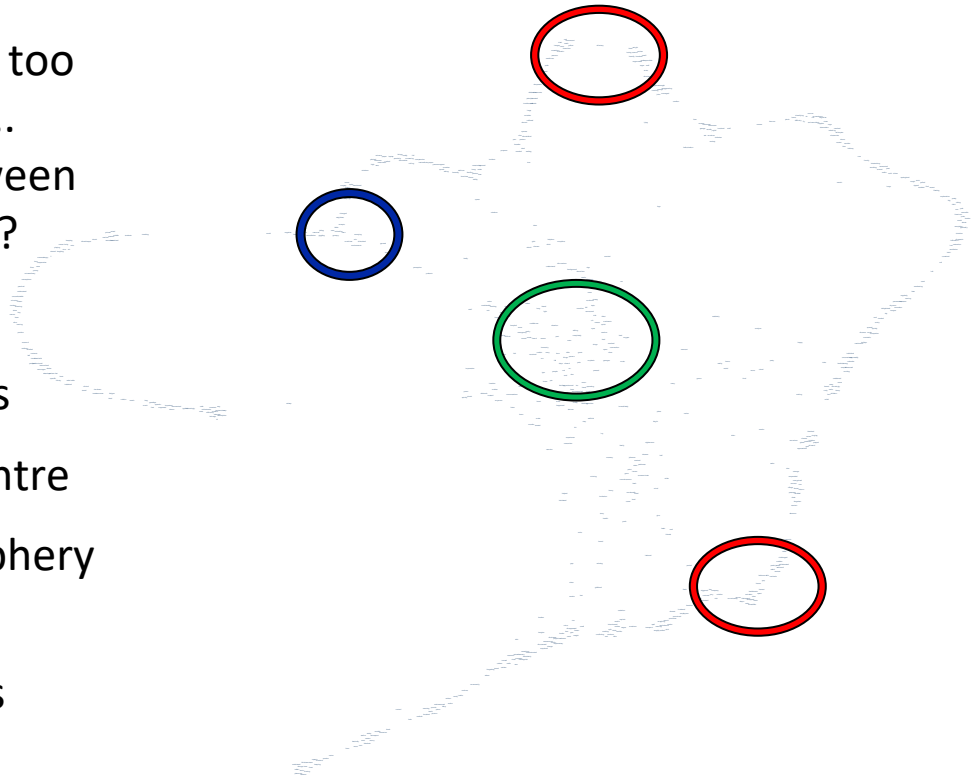
mein Kommentar: die Ebene an Reife ist beindruckend: Gelassenheit, leben mit der Krankheit. Sich konzentrieren auf das was das Leben ausmacht: Beziehungen zu Menschen, die postiven Dinge als lebenswert annehmen.

```
> closest_to(training_ms20,"koerper")
         word similarity to "koerper"
1      koerper              1.0000000
2       physik              0.5999420
3     koerpers              0.5681156
4         wahr              0.5043965
5      sondern              0.4641749
6  bewusstsein              0.4570825
7     staerken              0.4353815
8        power              0.4327700
9    benoetigt              0.4313358
10  intensiver              0.4302129
```

mein Kommentar: gegen die Physik kommt niemand an. Stärker wahrnehmen, intensiver erleben, das Bewusstsein stärken.

1.2 NLP Approaches to Analysis: Teaser 3: Conceptual Maps with DIPEx CMI

DIPEx CMI is only 8000 words, way too small for distributional semantics ... but can we explore the space between words and concepts, plot the plots?



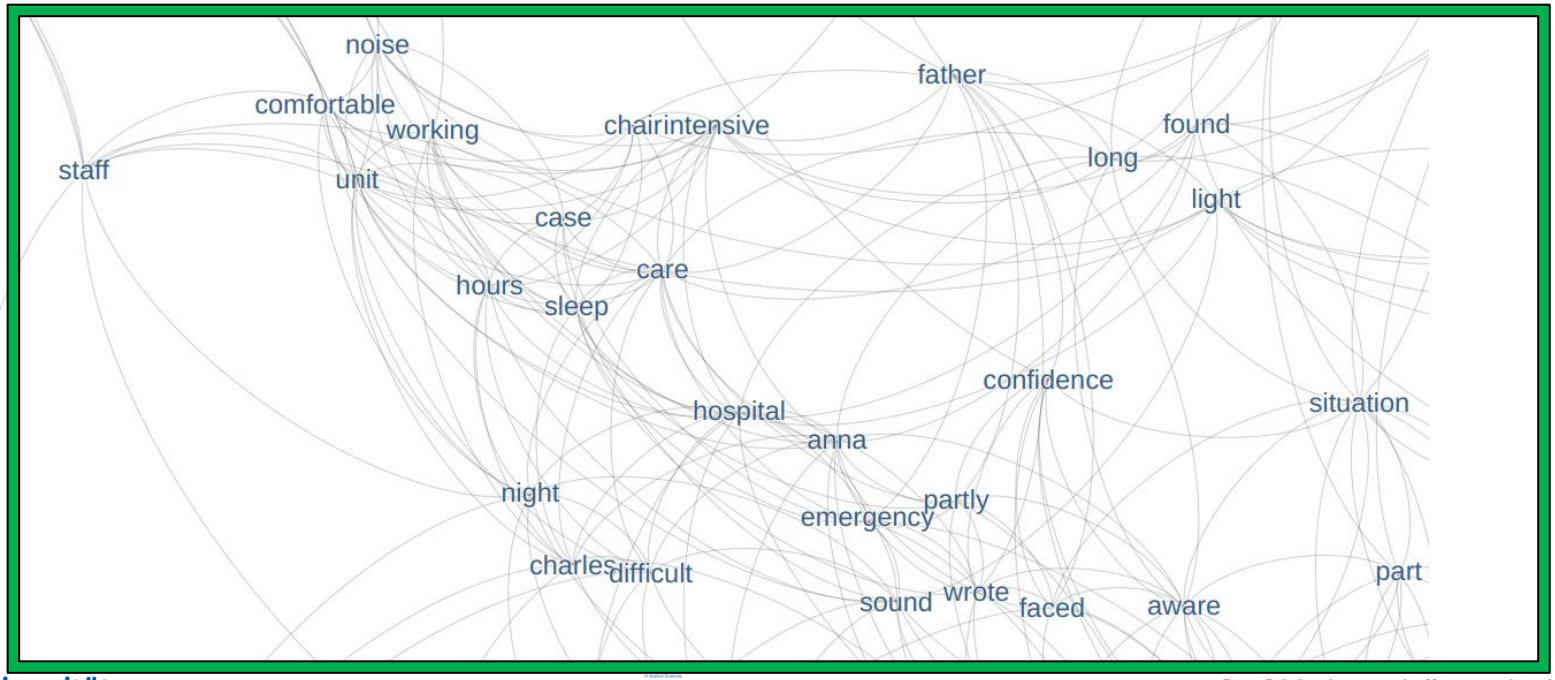
The conceptual map (500 w) shows

- Common core (green) in the centre
- Shared issues (blue) in the periphery
- Individual adventures (coma, hallucinations, red) in the edges

1.2 NLP Approaches to Analysis: Teaser 3: Conceptual Maps with DIPEx CMI & MS

The conceptual map (500 w) shows

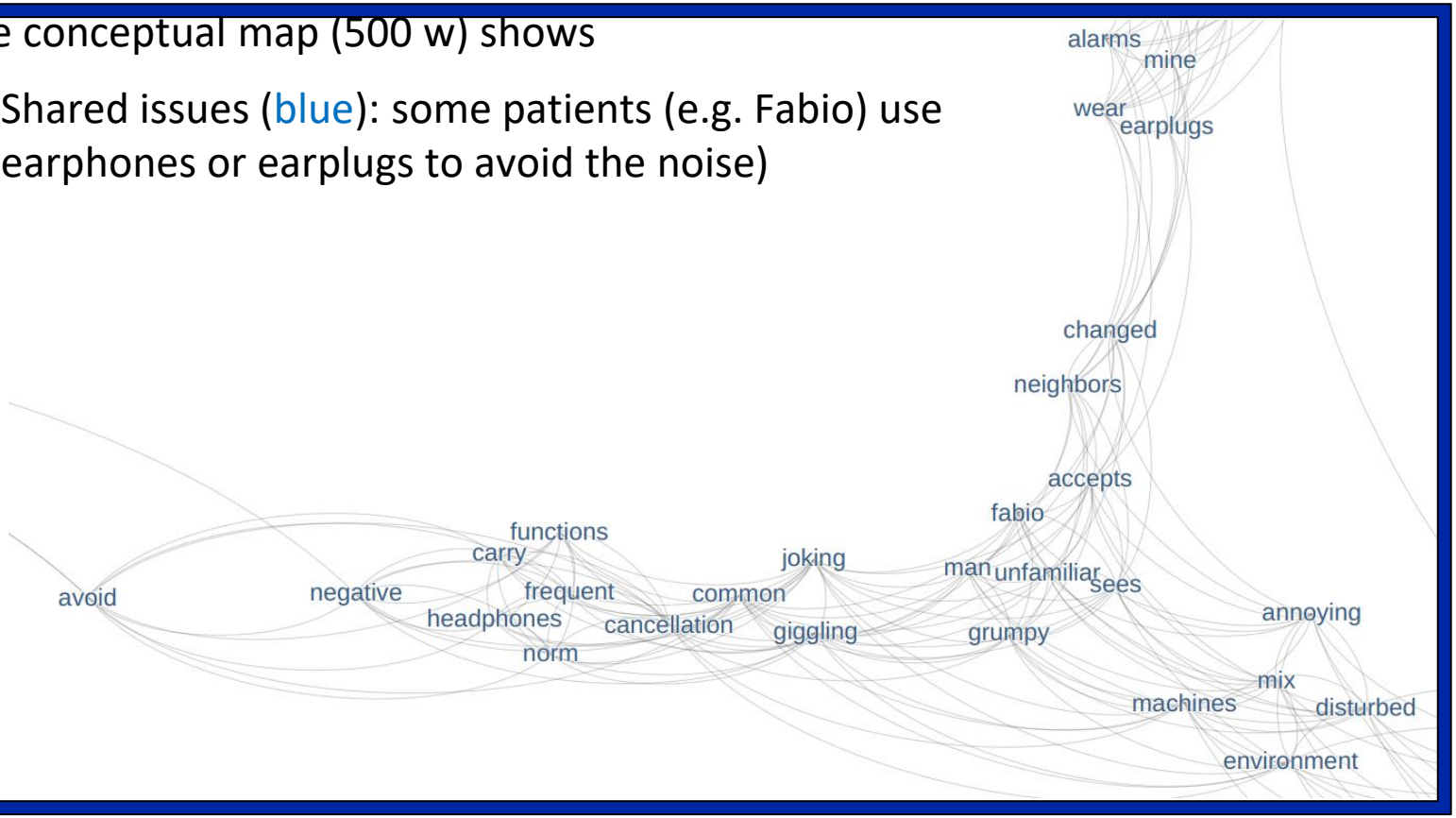
- Common semantic core (green): most patient (e.g. Charles or Anna) feel comfortable, but find it difficult to sleep, partly because of the noise



1.2 NLP Approaches to Analysis: Teaser 3: Conceptual Maps with DIPEx CMI & MS

The conceptual map (500 w) shows

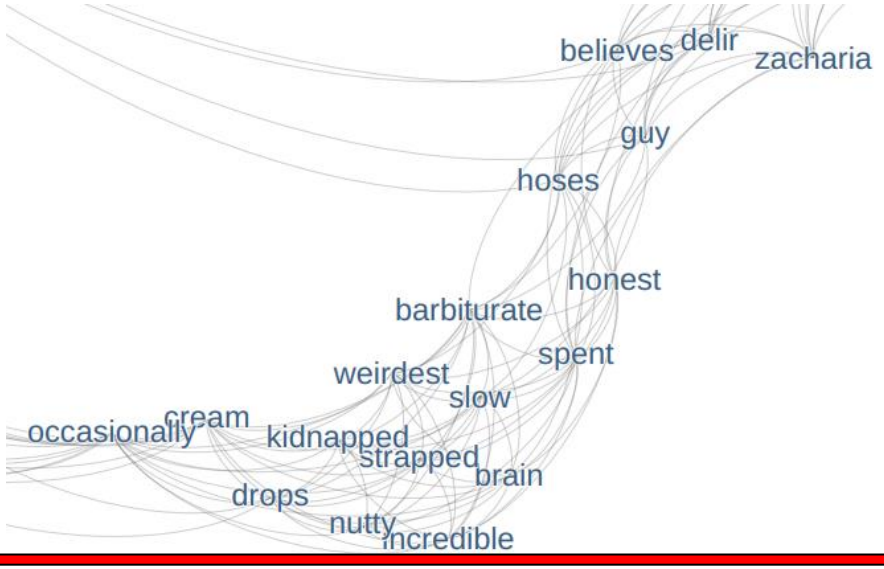
- Shared issues (blue): some patients (e.g. Fabio) use earphones or earplugs to avoid the noise)



1.2 NLP Approaches to Analysis: Teaser 3: Conceptual Maps with DIPEx CMI & MS

The conceptual map (500 w) shows

- Individual adventures (coma, hallucinations, red):
  Zacharia, in delir, believes that he is kidnapped.
  He finds that incredibly nutty, himself.



1.2 NLP Approaches to Analysis: Teaser 3: Conceptual Maps with DIPEx CMI & MS

From texts to maps … and back

"One of them was very funny, he had a curly beard, curly hair and was very, very sensitive, I noticed that … my friends who later came to visit me in the intensive care unit, they almost fell off their chairs when I first thanked them for being there and singing the song … my wife and children said stay here, stay here. And then I just prayed to God that He would give me another chance, that I would be allowed to return again"

1.2 NLP Approaches to Analysis: Teaser 1: Document Classification

Predict EXP_topic based on patient text and the description (EXP_description_ENG, EXP_textversion_ENG) with logistic regression.

Model Evaluation Met

Metric	Value
Accuracy	0.6829
Kappa	0.294

Model Confusion Matrix:

Act \ Pred	generalperecep...	stateofconsciu...
generalperece...	7	9
stateofconsciu...	4	21

Evaluation ^

Feature weights of class “state of consciousness”→

Feature	Frequency	Feature Infl...
tell	7	2.8189
coma	9	2.7887
dreams	4	2.7257
remembers	7	2.573
because	12	2.5301
<QUESTIONMA...	13	2.356
between	7	2.1189
/	9	2.0057
yeah	3	2.0006
gave	7	1.9681
later	7	1.9596
myself	6	1.912
back	7	1.8767
my	14	1.775
came	8	1.7002
dream	4	1.6395
ms.	2	1.6169
reality	6	1.6115

Conclusion

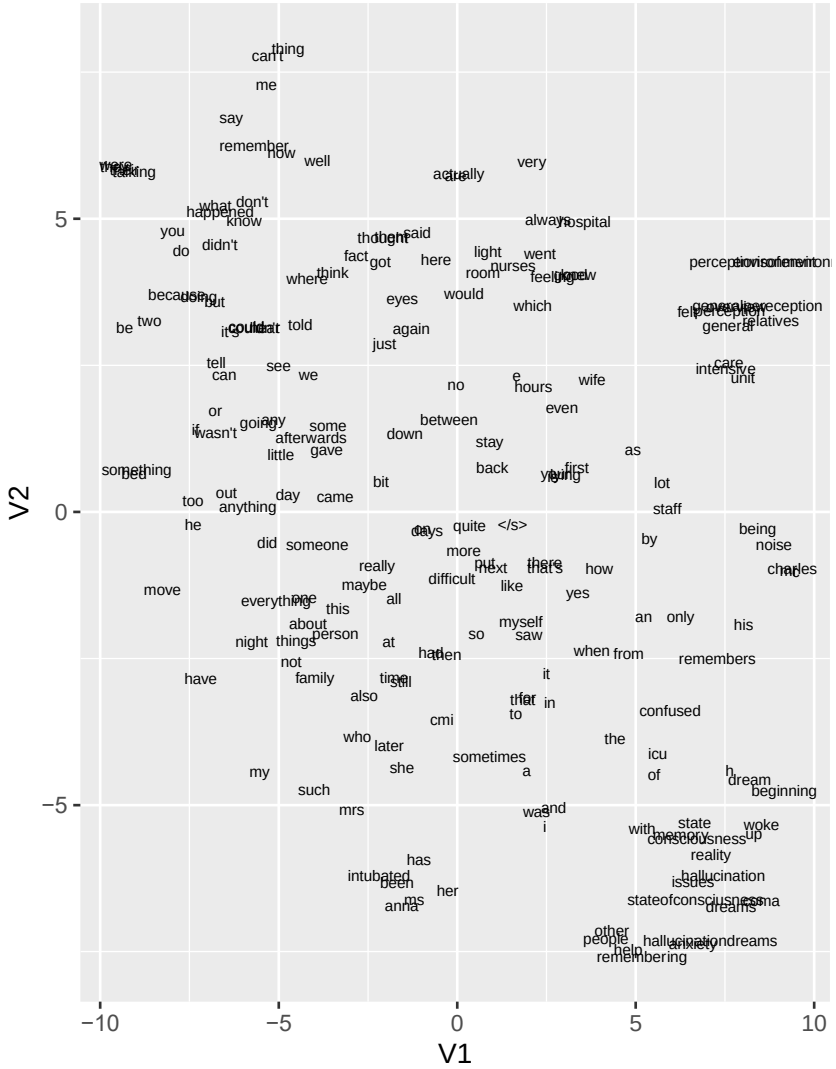
- Demonstrated the DIPEx DB and three pilot studies
- New insights are possible, but interpretation is important
- The abstraction from individual stories to concepts is not always smooth
- We need more transcribed texts (with or w/o metadata)
- Outlook: DSI Health workshops on medical chatbots

Leading questions for exchange

- Can we learn anything from these approaches?
- How can we evaluate and interpret the results?
- Due to sparseness, our approaches or only just about beginning to work. How can we get more texts?
- What further methods would you like to use?
- How can you imagine using these results in your work?

1.2 NLP Approaches to Analysis: Teaser 3: DIPEx CMI

DIPEx CMI is only 8000 words, too small for distributional semantics ...



DIPEx.ch
Health Experiences

Thanks for your attention!