[MCSQ]: Multilingual Corpus of Survey Questions

Dr. Diana Zavala-Rojas

SSHOC WP4 leader

Senior Research Fellow

Universitat Pompeu Fabra

ESS ERIC

Outline

- What is a corpus?
- Background
- The [MCSQ]: Multilingual Corpus of Survey Questionnaires
- Examples
- Design
- Access
- Q&A
-

What is a (text) corpus?

- **Linguistics**: a *corpus* (pl. *corpora*) is a large and structured set of texts (data). Texts are stored in machine readable format.
  - Are pdf's machine readable? = No.
  - Are MS Word documents machine readable? = No
- **Corpus linguistics**: uses corpus to conduct statistical analysis and hypothesis testing on language domains, checking occurrences or validating linguistic rules. Big data language apps use corpus data
- **Parallel corpus**: collection of aligned texts. Usually, alignments between a source text and its translations occur at the sentence level.
- **Famous corpus**:
  - DCEP: Digital Corpus of the European Parliament - European Commission documents (23 languages, 1.37 billion words, English segments 7.7 million (2000-2012)
  - Europarl Corpus consists of the proceedings of the European Parliament from 1996 (21 languages, 60 million words)
  - Oxford English corpus, the largest one: 2.1 billion words
  - Open subtitles: translated movie (and TV) subtitles

SSHOC

social sciences & humanities open cloud

Why a corpus of survey questionnaires?
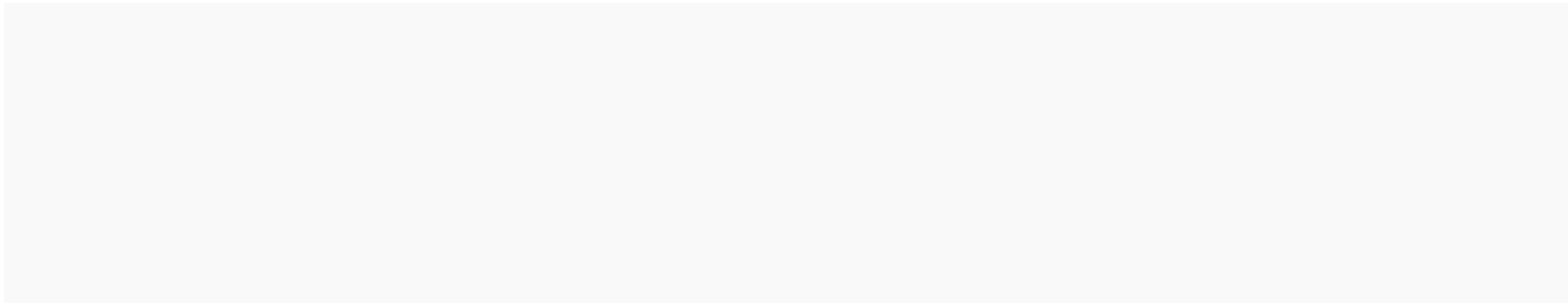
Background

- ESS, SHARE and EVS data archives provide similar information to end users: primary data in *.csv or similar formats and documentation including questionnaires in *.pdf format.

- **Challenge:** Access and use questions' texts for research purposes, systematically and in a comparative way.

- **E.g.**
  - To compare source items across rounds using the term 'democracy'.
  - To compare translations in French of a same question across languages
  - To get the questions' texts and copy them in the Appendix of an article to be submitted to a journal
  - To get questions in a format that allows programming of survey technologies that require survey questions texts in multiple languages (Qualtrics, electronic questionnaire device, etc.)
  - To implement technologies that require texts of diferent survey projects (Euro Panel)
  - To implement computational resources such as machine translation, prediction of question's complexity
  - To trace back breaks in the time series due to wording change
  - Etc....

SSHOC
social sciences & humanities open cloud

Background (II)

- **Accessing and comparing questionnaires texts manually can be challenging. In the ESS R8:**

A présent, nous aimerions vous poser quelques questions sur la politique et le gouvernement.

B1 Quel intérêt avez-vous pour la politique? Êtes-vous…?

LIRE

1: Très intéressé,
2: Assez intéressé,
3: Peu intéressé,
4: Ou, pas du tout intéressé
7: Refus
8: (Ne sait pas)

- **FRE- BE**

Base : at all

[Display : Et maintenant, quelques questions concernant la politique et le gouvernement.]

Base : at all

[S] **B1 - Vous intéressez-vous beaucoup, assez, peu ou pas du tout à la politique ?**

ENQUETEUR : CITER

Beaucoup	1
Assez	2
Peu	3
Pas du tout	4
(Refus)	5
(Ne sait pas)	6

- **FRE-FR**

- **Q108 B1**   *Exclusive*

Et maintenant quelques questions concernant la politique et le gouvernement.

B1 • Quel intérêt avez-vous pour la politique ? Êtes-vous …

_____

Enq. : LIRE

	1	… très intéressé,
	2	… assez intéressé,
	3	… peu intéressé,
	4	ou, pas du tout intéressé ?
7	5	-------> NE PAS LIRE : Refus
8	6	-------> NE PAS LIRE : Ne sait pas

Base : at all

- **FRE-CH**

SSHOC

social sciences & humanities open cloud

Background (III)

- **Accessing and comparing questionnaires texts manually can be challenging. In SHARE W7:**

IF *DN021_HighestEduPart* = a97
  Txt_FL_DN022

  DN022_OtherHighestEduPart
  Quel autre niveau de scolarisation ou diplôme de fin de scolarité [@ a/@ a/@ a/@ a/@ avait/@ avait] atteint votre
  [@ ex-/@ ex-/@ défunt/@ défunt(e)] [@ mari/@ femme/@ mari/@ femme/@ mari/@ femme] ?
  OTHER HIGHEST EDUCATIONAL DEGREE PARTNER OBTAINED
  STRING

ENDIF Txt_FL_DN023

- **FRE-BE**

IF *DN021_HighestEduPart* = a97
  Txt_FL_DN022

  DN022_OtherHighestEduPart
  Quel autre niveau d'études primaires ou secondaires a/avait atteint votre conjoint/ancien conjoint ?
  OTHER HIGHEST EDUCATIONAL DEGREE PARTNER OBTAINED
  STRING

ENDIF Txt_FL_DN023

- **FRE-FR**

IF *DN021_HighestEduPart* = a97
  Txt_FL_DN022

  DN022_OtherHighestEduPart
  Quel autre niveau de scolarité primaire ou secondaire [@ a/@ a/@ a/@ a/@ avait/@ avait] achevé votre [@ ex-/@
  ex-/@ (défunt)/@ (défunte)] [@ mari/@ femme/@ mari/@ femme/@ mari/@ femme] ?
  OTHER HIGHEST EDUCATIONAL DEGREE PARTNER OBTAINED
  STRING

ENDIF Txt_FL_DN023

- **FRE-CH**

Background (IV)

- **Accessing and comparing questionnaires texts manually can be challenging. In EVS 2008:**

Q2 Quand vous êtes entre amis, vous arrive-t-il souvent, de temps en temps ou jamais de discuter politique ?
(ENQ. NE RIEN SUGGERER - 1 SEULE REPONSE POSSIBLE)
1 – Souvent
2 – De temps en temps
3 – Jamais
8 – ne sait pas
9 – sans réponse

☐ (v7) **FRE-BE**

MONTREZ CARTE 2

Q2 Quand vous êtes entre amis, vous arrive-t-il souvent, de temps en temps ou jamais de discuter politique ?
1 – Souvent
2 – De temps en temps
3 – Jamais
8 – NSP (Spontané)
9 – SR (Spontané)

☐ (v7) • **FRE-FR**

MONTRER CARTE 3

Q2 Quand vous êtes entre amis, vous arrive-t-il souvent, de temps en temps ou jamais de discuter politique ?
1 – Souvent
2 – De temps en temps
3 – Jamais
8 – Ne sait pas (Spontané)
9 – Pas de réponse (Spontané)

☐ (v7) • **FRE-CH**

MONTRER CARTE 3

SSHOC
social sciences & humanities open cloud

Background (V)

- **Accessing and comparing questionnaires texts manually can be challenging. In EVS 1981:**

112. Comment préférez-vous passer votre temps libre ? (ENQ. UNE SEULE REPONSE)

- seul — 1
- en famille/ménage — 2
- avec des amis — 3
- dans un endroit animé, avec pas mal de monde — 4
- tout également — 5
- ne sait pas — Y

FRE-BE

Q.112 – Et préférez-vous passer votre temps libre ...

1 seule réponse

... seul ? 1
... en famille ? 2
... avec des amis ? 3
... ou dans un endroit animé avec beaucoup de monde ? . 4
tout également 5
N.S.P. ..... Y

FRE-FR

112. And during your leisure time do you prefer to be alone, to be with your family, to be with friends or to be in a lively place with many people?

ALONE----1
WITH FAMILY----2
WITH FRIENDS----3
IN A LIVELY PLACE----4
ALL EQUALLY----5
DON'T KNOW----Y

1+2 = 6
1+3 = 7
1+4 = 8
2+3 = 9
2+4 = 0
3+4 = X

ENG-CA

- In summary, end users have easy access to primary data, but usability of the documentation, including questionnaires, is not well developed yet (Mohler et al 2010).

Solution: a database (corpus) of survey questionnaires

- Provides machine readable data that has valuable metadata associated to it, which is not present in PDF questionnaires, to identify clearly the questions, their linguistic components and the survey projects.

Corpus vs. question bank

- A corpus can be considered a question bank, because it is a collection of questions and metadata.

- A question bank cannot be considered a corpus, because it is not necessarily designed as linguistic data

SSHOC

social sciences & humanities open cloud

[MCSQ]: some examples from the (prototype) interface)

http://easy.mcsq.upf.edu/

SSHOC

social sciences & humanities open cloud

- Compare the "agree" items in French in France and Belgium in ESSR6

MCSQ

Comparing by word

	survey_itemid	Text	item_name	item_type	survey_itemid	Text	item_name	item_type
0	ESS_R06_2012_FRE_FR_273	Pour chacune d'elles, dites-moi si vous êtes d'accord ou pas d'accord.	B26	REQUEST	ESS_R06_2012_FRE_BE_255	À l'aide de cette carte, dites-moi s'il vous plaît dans quelle mesure vous êtes d'accord ou non avec chacune des propositions suivantes?	B26	REQUEST
1	ESS_R06_2012_FRE_FR_276	TOUT A FAIT D'ACCORD	B26	RESPONSE	ESS_R06_2012_FRE_BE_258	Tout à fait d'accord	B26	RESPONSE
2	ESS_R06_2012_FRE_FR_277	PLUTOT D'ACCORD	B26	RESPONSE	ESS_R06_2012_FRE_BE_259	Plutôt d'accord	B26	RESPONSE
3	ESS_R06_2012_FRE_FR_278	NI D'ACCORD, NI PAS D'ACCORD	B26	RESPONSE	ESS_R06_2012_FRE_BE_260	Ni d'accord, ni en désaccord	B26	RESPONSE
4	ESS_R06_2012_FRE_FR_279	PLUTOT PAS D'ACCORD	B26	RESPONSE	ESS_R06_2012_FRE_BE_261	Plutôt en désaccord	B26	RESPONSE
5	ESS_R06_2012_FRE_FR_280	PAS DU TOUT D'ACCORD	B26	RESPONSE	ESS_R06_2012_FRE_BE_262	Tout à fait en désaccord	B26	RESPONSE
0	ESS_R06_2012_FRE_FR_284	TOUT A FAIT D'ACCORD	B27	RESPONSE	ESS_R06_2012_FRE_BE_266	Tout à fait d'accord	B27	RESPONSE
1	ESS_R06_2012_FRE_FR_285	PLUTOT D'ACCORD	B27	RESPONSE	ESS_R06_2012_FRE_BE_267	Plutôt d'accord	B27	RESPONSE
2	ESS_R06_2012_FRE_FR_286	NI D'ACCORD, NI PAS D'ACCORD	B27	RESPONSE	ESS_R06_2012_FRE_BE_268	Ni d'accord, ni en désaccord	B27	RESPONSE
3	ESS_R06_2012_FRE_FR_287	PLUTOT PAS D'ACCORD	B27	RESPONSE	ESS_R06_2012_FRE_BE_269	Plutôt en désaccord	B27	RESPONSE
4	ESS_R06_2012_FRE_FR_288	PAS DU TOUT D'ACCORD	B27	RESPONSE	ESS_R06_2012_FRE_BE_270	Tout à fait en désaccord	B27	RESPONSE
0	ESS_R06_2012_FRE_FR_679	En utilisant cette carte, veuillez me dire dans quelle mesure vous êtes d'accord ou non avec les propositions suivantes.	D16	REQUEST	ESS_R06_2012_FRE_BE_663	En utilisant cette carte, pouvez-vous me dire dans quelle mesure vous êtes d'accord ou non avec les propositions suivantes?	D16	REQUEST
1	ESS_R06_2012_FRE_FR_681	TOUT A FAIT D'ACCORD	D16	RESPONSE	ESS_R06_2012_FRE_BE_666	Tout à fait d'accord	D16	RESPONSE
2	ESS_R06_2012_FRE_FR_682	PLUTOT D'ACCORD	D16	RESPONSE	ESS_R06_2012_FRE_BE_667	Plutôt d'accord	D16	RESPONSE
3	ESS_R06_2012_FRE_FR_683	NI D'ACCORD, NI PAS D'ACCORD	D16	RESPONSE	ESS_R06_2012_FRE_BE_668	Ni d'accord, ni en désaccord	D16	RESPONSE
4	ESS_R06_2012_FRE_FR_684	PLUTOT PAS D'ACCORD	D16	RESPONSE	ESS_R06_2012_FRE_BE_669	Plutôt en désaccord	D16	RESPONSE
5	ESS_R06_2012_FRE_FR_685	PAS DU TOUT D'ACCORD	D16	RESPONSE	ESS_R06_2012_FRE_BE_670	Tout en fait en désaccord	D16	RESPONSE
0	ESS_R06_2012_FRE_FR_689	TOUT A FAIT D'ACCORD	D17	RESPONSE	ESS_R06_2012_FRE_BE_674	Tout à fait d'accord	D17	RESPONSE
1	ESS_R06_2012_FRE_FR_690	PLUTOT D'ACCORD	D17	RESPONSE	ESS_R06_2012_FRE_BE_675	Plutôt d'accord	D17	RESPONSE
2	ESS_R06_2012_FRE_FR_691	NI D'ACCORD, NI PAS D'ACCORD	D17	RESPONSE	ESS_R06_2012_FRE_BE_676	Ni d'accord, ni en désaccord	D17	RESPONSE
3	ESS_R06_2012_FRE_FR_692	PLUTOT PAS D'ACCORD	D17	RESPONSE	ESS_R06_2012_FRE_BE_677	Plutôt en désaccord	D17	RESPONSE
4	ESS_R06_2012_FRE_FR_693	PAS DU TOUT D'ACCORD	D17	RESPONSE	ESS_R06_2012_FRE_BE_678	Tout en fait en désaccord	D17	RESPONSE
0	ESS_R06_2012_FRE_FR_697	TOUT A FAIT D'ACCORD	D18	RESPONSE	ESS_R06_2012_FRE_BE_682	Tout à fait d'accord	D18	RESPONSE
1	ESS_R06_2012_FRE_FR_698	PLUTOT D'ACCORD	D18	RESPONSE	ESS_R06_2012_FRE_BE_683	Plutôt d'accord	D18	RESPONSE
2	ESS_R06_2012_FRE_FR_699	NI D'ACCORD, NI PAS D'ACCORD	D18	RESPONSE	ESS_R06_2012_FRE_BE_684	Ni d'accord, ni en désaccord	D18	RESPONSE
3	ESS_R06_2012_FRE_FR_700	PLUTOT PAS D'ACCORD	D18	RESPONSE	ESS_R06_2012_FRE_BE_685	Plutôt en désaccord	D18	RESPONSE
4	ESS_R06_2012_FRE_FR_701	PAS DU TOUT D'ACCORD	D18	RESPONSE	ESS_R06_2012_FRE_BE_686	Tout en fait en désaccord	D18	RESPONSE
0	ESS_R06_2012_FRE_FR_705	TOUT A FAIT D'ACCORD	D19	RESPONSE	ESS_R06_2012_FRE_BE_690	Tout à fait d'accord	D19	RESPONSE

social sciences & humanities open cloud

| 0 | ESS_R06_2012_FRE_BE_663 | En utilisant cette carte, pouvez-vous me dire dans quelle mesure vous êtes d'accord ou non avec les propositions suivantes? | D16 | REQUEST |
| | ESS_R06_2012_FRE_FR_679 | En utilisant cette carte, veuillez me dire dans quelle mesure vous êtes d'accord ou non avec les propositions suivantes. | D16 | REQUEST |

SSHOC
social sciences & humanities open cloud

- **Compare EVS response scales from 2008 in German in Germany and Switzerland, and include the French Swiss version**

MCSQ

MCSQ Official Site | Home | User Guidelines | Changelogs | Hi, Diana Zavala-Rojas | Compare versions | Word search in MCSQ Tables | MCSQ Tables

Compare by item type

Select study **i** | EVS

Select year **i** | 2008

Select item type **i** | RESPONSE

Select language/country 1 **i** | GER_DE

Select language/country 2 **i** | GER_CH

Select language/country 3 **i** | FRE_CH

Select language/country 4 **i** | No filter

Select language/country 5 **i** | No filter

Select language/country 6 **i** | No filter

Select language/country 7 **i** | No filter

Select language/country 8 **i** | No filter

Submit

[MCSQ]: The Multilingual Corpus of Survey Questionnaires

MCSQ

Comparing by item type

	survey_itemid	Text	item_name	item_type	survey_itemid	Text	item_name	item_type	survey_itemid	Text	item_name	item_type
0	EVS_R04_2008_GER_DE_352	Überhaupt keine Freiheit	Q10	RESPONSE	EVS_R04_2008_GER_CH_352	Überhaupt keine Freiheit	Q10	RESPONSE	EVS_R04_2008_FRE_CH_356	Pas du tout libre	Q10	RESPONSE
1	EVS_R04_2008_GER_DE_353	Völlige Freiheit	Q10	RESPONSE	EVS_R04_2008_GER_CH_353	Völlige Freiheit	Q10	RESPONSE	EVS_R04_2008_FRE_CH_357	Tout à fait libre	Q10	RESPONSE
2	EVS_R04_2008_GER_DE_362	Weiß nicht	Q10	RESPONSE	EVS_R04_2008_GER_CH_362	Weiss nicht	Q10	RESPONSE	EVS_R04_2008_FRE_CH_366	Ne sait pas	Q10	RESPONSE
3	EVS_R04_2008_GER_DE_363	Verweigert	Q10	RESPONSE	EVS_R04_2008_GER_CH_363	Keine Antwort	Q10	RESPONSE	EVS_R04_2008_FRE_CH_367	Pas de réponse	Q10	RESPONSE
0	EVS_R04_2008_GER_DE_2353	Ja	Q100	RESPONSE	EVS_R04_2008_GER_CH_2313	Ja	Q100	RESPONSE	EVS_R04_2008_FRE_CH_2283	Oui	Q100	RESPONSE
1	EVS_R04_2008_GER_DE_2354	Nein	Q100	RESPONSE	EVS_R04_2008_GER_CH_2314	Nein	Q100	RESPONSE	EVS_R04_2008_FRE_CH_2284	Non	Q100	RESPONSE
2	EVS_R04_2008_GER_DE_2355	Trifft nicht zu	Q100	RESPONSE	EVS_R04_2008_GER_CH_2315	Trifft nicht zu	Q100	RESPONSE	EVS_R04_2008_FRE_CH_2285	Non applicable	Q100	RESPONSE
3	EVS_R04_2008_GER_DE_2356	Weiß nicht (spontan)	Q100	RESPONSE	EVS_R04_2008_GER_CH_2316	weiss nicht (spontan)	Q100	RESPONSE	EVS_R04_2008_FRE_CH_2286	Ne sait pas (Spontané, ne rien suggérer)	Q100	RESPONSE
4	EVS_R04_2008_GER_DE_2357	Antwort verweigert (spontan)	Q100	RESPONSE	EVS_R04_2008_GER_CH_2317	keine Antwort (spontan)	Q100	RESPONSE	EVS_R04_2008_FRE_CH_2287	Pas de réponse (Spontané, ne rien suggérer)	Q100	RESPONSE
0	EVS_R04_2008_GER_DE_2362	Ja	Q101	RESPONSE	EVS_R04_2008_GER_CH_2323	Ja	Q101	RESPONSE	EVS_R04_2008_FRE_CH_2293	Oui	Q101	RESPONSE
1	EVS_R04_2008_GER_DE_2363	Nein	Q101	RESPONSE	EVS_R04_2008_GER_CH_2324	Nein	Q101	RESPONSE	EVS_R04_2008_FRE_CH_2294	Non	Q101	RESPONSE
2	EVS_R04_2008_GER_DE_2364	Trifft nicht zu	Q101	RESPONSE	EVS_R04_2008_GER_CH_2325	Trifft nicht zu	Q101	RESPONSE	EVS_R04_2008_FRE_CH_2295	Non applicable	Q101	RESPONSE
3	EVS_R04_2008_GER_DE_2365	Weiß nicht (spontan)	Q101	RESPONSE	EVS_R04_2008_GER_CH_2326	weiss nicht (spontan)	Q101	RESPONSE	EVS_R04_2008_FRE_CH_2296	Ne sait pas (Spontané, ne rien suggérer)	Q101	RESPONSE
4	EVS_R04_2008_GER_DE_2366	Antwort verweigert (spontan)	Q101	RESPONSE	EVS_R04_2008_GER_CH_2327	keine Antwort (spontan)	Q101	RESPONSE	EVS_R04_2008_FRE_CH_2297	Pas de réponse (Spontané, ne rien suggérer)	Q101	RESPONSE
0	EVS_R04_2008_GER_DE_2369	Ja, verheiratet	Q102	RESPONSE	EVS_R04_2008_GER_CH_2331	ja, verheiratet	Q102	RESPONSE	EVS_R04_2008_FRE_CH_2301	Oui, marié	Q102	RESPONSE
1	EVS_R04_2008_GER_DE_2370	Ja, eingetragene Lebenspartnerschaft	Q102	RESPONSE	EVS_R04_2008_GER_CH_2332	ja, registrierte Partnerschaft	Q102	RESPONSE	EVS_R04_2008_FRE_CH_2302	Oui, en partenariat enregistré	Q102	RESPONSE
2	EVS_R04_2008_GER_DE_2371	Nein	Q102	RESPONSE	EVS_R04_2008_GER_CH_2333	Nein	Q102	RESPONSE	EVS_R04_2008_FRE_CH_2303	Non	Q102	RESPONSE
3	EVS_R04_2008_GER_DE_2372	Trifft nicht zu	Q102	RESPONSE	EVS_R04_2008_GER_CH_2334	Trifft nicht zu	Q102	RESPONSE	EVS_R04_2008_FRE_CH_2304	Non applicable	Q102	RESPONSE
4	EVS_R04_2008_GER_DE_2373	Weiß nicht (spontan)	Q102	RESPONSE	EVS_R04_2008_GER_CH_2335	weiss nicht (spontan)	Q102	RESPONSE	EVS_R04_2008_FRE_CH_2305	Ne sait pas (Spontané, ne rien suggérer)	Q102	RESPONSE
5	EVS_R04_2008_GER_DE_2374	Antwort verweigert (spontan)	Q102	RESPONSE	EVS_R04_2008_GER_CH_2336	keine Antwort (spontan)	Q102	RESPONSE	EVS_R04_2008_FRE_CH_2306	Pas de réponse (Spontané, ne rien suggérer)	Q102	RESPONSE
0	EVS_R04_2008_GER_DE_2376	Ja	Q103	RESPONSE	EVS_R04_2008_GER_CH_2338	Ja	Q103	RESPONSE	EVS_R04_2008_FRE_CH_2308	Oui	Q103	RESPONSE
1	EVS_R04_2008_GER_DE_2377	Nein	Q103	RESPONSE	EVS_R04_2008_GER_CH_2339	Nein	Q103	RESPONSE	EVS_R04_2008_FRE_CH_2309	Non	Q103	RESPONSE

SSHOC
social sciences & humanities open cloud

- **Compare source questions using the expression "income"**

MCSQ

Search results for the word income in MCSQ Request Collection

	survey_itemid	Text	item_name	country_language	moduleid
0	ESS_R02_2004_ENG_SOURCE_331	The government should take measures to reduce differences in income levels	B30	ENG_SOURCE	2
1	ESS_R02_2004_ENG_SOURCE_1515	Please consider the income of all household members and any income which may be received by the household as a whole.	F31	ENG_SOURCE	6
2	ESS_R02_2004_ENG_SOURCE_1516	What is the main source of income in your household?	F31	ENG_SOURCE	6
3	ESS_R02_2004_ENG_SOURCE_1529	Using this card, if you add up the income from all sources, which letter describes your household's total net income?	F32w	ENG_SOURCE	6
4	ESS_R02_2004_ENG_SOURCE_1531	Use the part of the card that you know best: weekly, monthly or annual income.	F32w	ENG_SOURCE	6
5	ESS_R02_2004_ENG_SOURCE_1548	Using this card, if you add up the income from all sources, which letter describes your household's total net income?	F32m	ENG_SOURCE	6
6	ESS_R02_2004_ENG_SOURCE_1550	Use the part of the card that you know best: weekly, monthly or annual income.	F32m	ENG_SOURCE	6
7	ESS_R02_2004_ENG_SOURCE_1567	Using this card, if you add up the income from all sources, which letter describes your household's total net income?	F32y	ENG_SOURCE	6
8	ESS_R02_2004_ENG_SOURCE_1569	Use the part of the card that you know best: weekly, monthly or annual income.	F32y	ENG_SOURCE	6
9	ESS_R02_2004_ENG_SOURCE_1586	Around how large a proportion of the household income do you provide yourself?	F32a	ENG_SOURCE	6
10	ESS_R02_2004_ENG_SOURCE_1597	Which of the descriptions on this card comes closest to how you feel about your household's income nowadays?	F33	ENG_SOURCE	6
11	ESS_R02_2004_ENG_SOURCE_2722	A high income	G112	ENG_SOURCE	23
12	ESS_R03_2006_ENG_SOURCE_329	The government should take measures to reduce differences in income levels	B30	ENG_SOURCE	2
13	ESS_R03_2006_ENG_SOURCE_1985	Please consider the income of all household members and any income which may be received by the household as a whole.	F31	ENG_SOURCE	6
14	ESS_R03_2006_ENG_SOURCE_1309	How worried are you, if at all, that your income in old age will not be adequate enough to cover your later years.	D53	ENG_SOURCE	4
15	ESS_R03_2006_ENG_SOURCE_1739	How important is it for you to compare your income with other people's incomes?	E54	ENG_SOURCE	5
16	ESS_R03_2006_ENG_SOURCE_1747	Whose income would you be most likely to compare your own with?	E55	ENG_SOURCE	5
17	ESS_R03_2006_ENG_SOURCE_1986	What is the main source of income in your household?	F31	ENG_SOURCE	6
18	ESS_R03_2006_ENG_SOURCE_1999	Using this card, if you add up the income from all sources, which letter describes your household's total net income?	F32w	ENG_SOURCE	6

- **Map how 'agree' has been translated in Russian across countries**

[MCSQ]: The Multilingual Corpus of Survey Questionnaires

SSHOC
social sciences & humanities open cloud

MCSQ

Search results retrieved from MCSQ Alignment Collection

	source_survey_itemid	target_survey_itemid	source_text	target_text
0	ESS_R06_2012_ENG_SOURCE_241	ESS_R06_2012_RUS_EE_258	Using this card, please say to what extent you agree or disagree with each of the following statements.	Скажите пожалуйста, насколько Вы согласны или не согласны со следующими утверждениями.
1	ESS_R06_2012_ENG_SOURCE_246	ESS_R06_2012_RUS_EE_264	Neither agree nor disagree	Ни согласен(на), ни против этого
2	ESS_R06_2012_ENG_SOURCE_247	ESS_R06_2012_RUS_EE_265	Disagree	Не согласен(на)
3	ESS_R06_2012_ENG_SOURCE_248	ESS_R06_2012_RUS_EE_266	Disagree strongly	Совершенно не согласен(на)
4	ESS_R06_2012_ENG_SOURCE_254	ESS_R06_2012_RUS_EE_272	Neither agree nor disagree	Ни согласен(на), ни против этого
5	ESS_R06_2012_ENG_SOURCE_255	ESS_R06_2012_RUS_EE_273	Disagree	Не согласен(на)
6	ESS_R06_2012_ENG_SOURCE_256	ESS_R06_2012_RUS_EE_274	Disagree strongly	Совершенно не согласен(на)
7	ESS_R06_2012_ENG_SOURCE_551	ESS_R06_2012_RUS_EE_572	Using this card, please say how much you agree or disagree with each of the following statements.	Скажите пожалуйста, насколько Вы согласны или не согласны со следующими утверждениями.
8	ESS_R06_2012_ENG_SOURCE_556	ESS_R06_2012_RUS_EE_578	Neither agree nor disagree	Ни согласен(на), ни против этого
9	ESS_R06_2012_ENG_SOURCE_557	ESS_R06_2012_RUS_EE_579	Disagree	Ни согласен(на)
10	ESS_R06_2012_ENG_SOURCE_558	ESS_R06_2012_RUS_EE_580	Disagree strongly	Совершенно несогласен(на)
11	ESS_R06_2012_ENG_SOURCE_564	ESS_R06_2012_RUS_EE_586	Neither agree nor disagree	Ни согласен(на), ни против этого
12	ESS_R06_2012_ENG_SOURCE_565	ESS_R06_2012_RUS_EE_587	Disagree	Ни согласен(на)
13	ESS_R06_2012_ENG_SOURCE_566	ESS_R06_2012_RUS_EE_588	Disagree strongly	Совершенно несогласен(на)
14	ESS_R06_2012_ENG_SOURCE_572	ESS_R06_2012_RUS_EE_594	Neither agree nor disagree	Ни согласен(на), ни против этого
15	ESS_R06_2012_ENG_SOURCE_573	ESS_R06_2012_RUS_EE_595	Disagree	Ни согласен(на)
16	ESS_R06_2012_ENG_SOURCE_574	ESS_R06_2012_RUS_EE_596	Disagree strongly	Совершенно несогласен(на)
17	ESS_R06_2012_ENG_SOURCE_661	ESS_R06_2012_RUS_EE_683	Using this card, please tell me to what extent you agree or disagree with each of the following statements.	Скажите пожалуйста, насколько Вы согласны или не согласны со следующими утверждениями.

social sciences & humanities open cloud

The MCSQ: Multilngual corpus of survey questionnaires

- **MCSQ**: a large and structured collection of questionnaires texts implemented as an Entity-Relationship database in SQL.
- It is open-access and open source. In line with the FAIR data principles.

- Version 1.1 (Beatrice Worsley):
- 169 distinct questionnaires (ESS, EVS and SHARE Covid)
- Largest coverage for ESS
- Data:
  – Nine languages: English source and its translations into Catalan, Czech, French, German, Norwegian, Portuguese, Spanish and Russian
  – More than 20 countries
  – Largest subset Russian: Azerbaijan, Belarus, Estonia, Israel, Latvia, Lithuania, Moldova, Russia, Ukraine
  – 404,041 segments (sentences and shorter linguistic units)

- Nomenclature of the data provides information to identify survey items:

SSHOC
social sciences & humanities open cloud

The MCSQ: behind the scenes
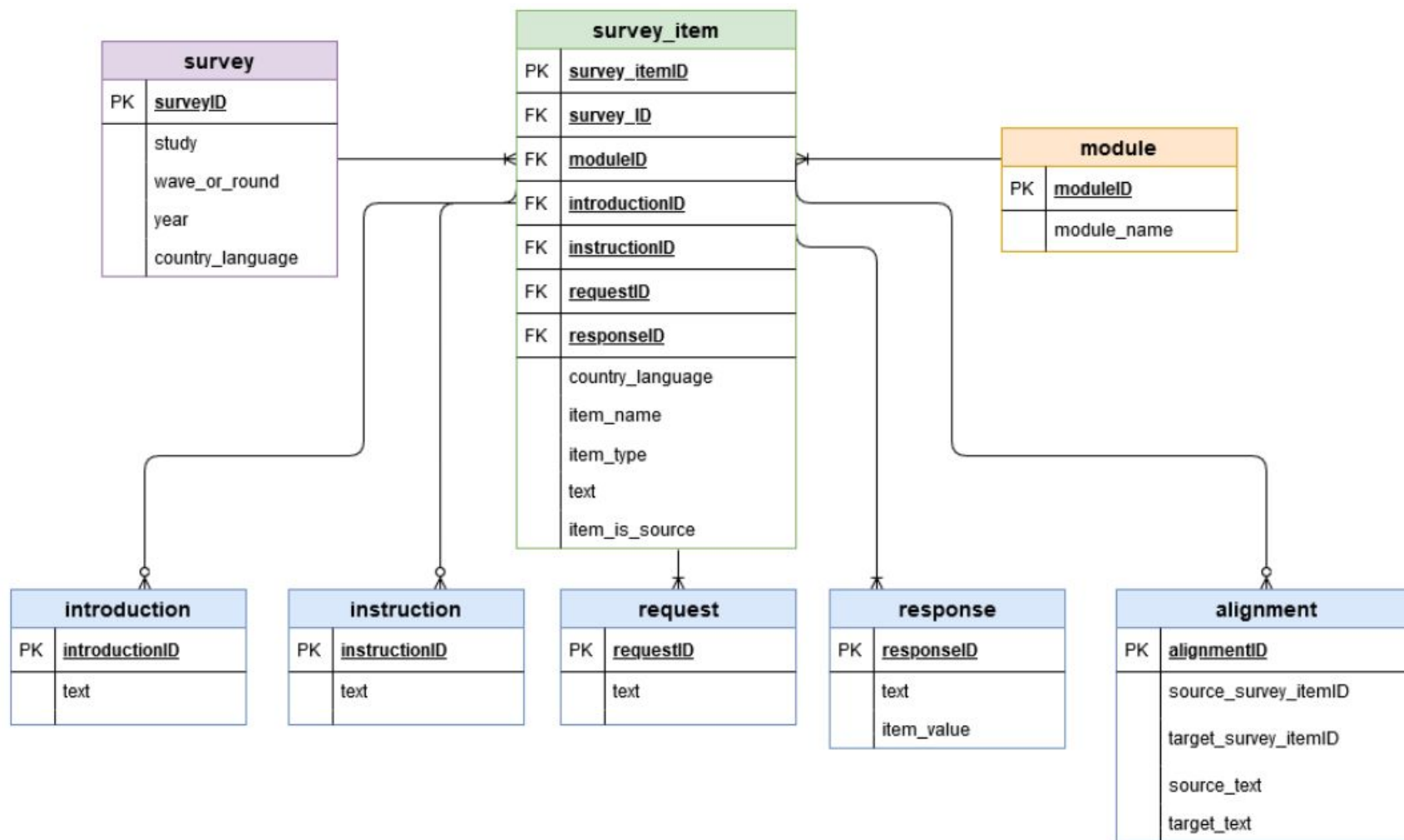
The MCSQ: behind the scenes

- The **MCSQ** decomposes survey items into four components: introduction, instruction, request and response**.**

- This classification allowed to develop a tailored alignment method

- The corpus is designed to be scalable, new data is added easily

- Access:
  – Dedicated website:
  https://www.upf.edu/web/mcsq

  – PostgreSQL platform:
  http://mcsq.upf.edu/login
  Write us to have access keys!

  – (prototype) interface
  http://easy.mcsq.upf.edu/

The MCSQ: Multilngual corpus of survey questionnaires

- Add more data

- Refine queries to the users' interface

- Add some functionalities (e.g. customized downloadable data)

- Add linguistic annotations to the data

- Test the corpus in the context of translation procedures (translation memory)

- Dissemination and publishing:

Zavala-Rojas, D., Sorato, D., Hareide, L., & Hofland, K. (forthcoming 2021). [MCSQ] Multilingual Corpus of Survey Questionnaires. *META: Journal Des Traducteurs.*

- Refine the (theoretical) TRAPD model with the use of corpus linguistics

SSHOC

social sciences & humanities open cloud

Conclusions

- Access to primary data is well developed across RIs
- Access to secondary data, including questionnaires' text is still underdeveloped
- The [MCSQ] aims to contribute to users' experience by making survey questions easily available
- It will support programming of survey tools by providing the texts in machine readable format
- The [MCSQ] provides a platform to improve translation procedures, documentation and research by allowing search and comparison of questions' texts across languages, within languages and across survey projects
- Finally, it promotes linguistic analysis of survey questionnaires e.g text analysis and comparison of questionnaires, search for neighbour terms (collocations) by storing texts as data.
- Explore it at https://www.upf.edu/web/mcsq

SSHOC
social sciences & humanities open cloud

Thank you for your attention!

https://www.upf.edu/web/mcsq

✉ **Diana.zavala@upf.edu**

🌐 **https://www.sshopencloud.eu**

✉ **info@sshopencloud.eu**

🐦 **@SSHOpenCloud**

💼 **/in/sshopencloud**

SSHOC
social sciences & humanities open cloud
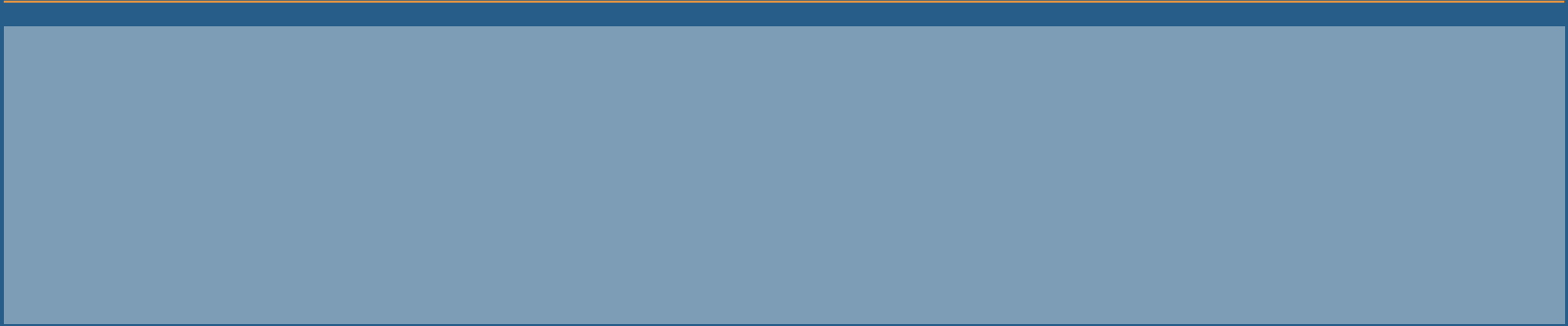
Click to edit Master title style