Tenacious parallel protocorpora and resulting lexicogrammatical resources: Henri Frei’s *Le livre des deux mille phrases*

Anthony Grant
Edge Hill University
granta@edgehill.ac.uk
Purpose

This paper examines the uses of a family of interlinked pre-modern parallel corpora (“protocorpora”), and focuses on their compilation and their use to derive what can be described as early lexicogrammatical resources.

These materials are pre-modern inasmuch as they date from the pre-transistorised era, and were therefore not assembled or stored electronically, but have been used, amplified and analysed since the 1940s.
Instances of protocorpora

- Collections of disparate texts in a common language, for instance the *Corpus Inscriptionum Latinarum*, a collection of Latin inscriptions from many times and places: [cil.bbaw.de](http://cil.bbaw.de)

- Translations of a pre-existing text, for instance the fourth century Arian Bible translation from Greek into Gothic by Wulfila

- As the first example shows, protocorpora can be turned into corpora.
- So can the second: [http://www.ub.uu.se/about-the-library/exhibitions/codex-argenteus/about-the-project/](http://www.ub.uu.se/about-the-library/exhibitions/codex-argenteus/about-the-project/)

- The “Frei family” protocorpora discussed here resemble the former, as they comprise collections of discrete sentences without overt textual cohesiveness.
In 1953 the Swiss Saussurean linguist Henri Frei published material from a corpus of spoken French which he had collected through the use of stimuli from Jean Barbéry (b. 1931), an educated speaker of Parisian French. The material which Frei collected was published in the form of 2000 single sentences, or as pairs of questions and responses (Frei, 1953, revised 1966): *Le Livre des deux mille phrases*.

Frei referred to these utterances as ‘phrases’; rather than as ‘énoncés’, the normal French translation of ‘sentence’, which is what they are: major or minor sentences.
This material, and the approach employed, were not new (though Frei’s account of exactly how he collected his material is nebulous).

Frei had already collected parallel protocorpus material (involving the same sentence cues) from speakers of Mandarin Chinese and from Japanese during his stay in the Far East (1932-1939).

He published a little of the former in Frei (1940-41), as sample sentences for an article in which he interrogated the material in an investigation of Mandarin tense and aspect.
Henri Frei (June 5 1899 – Nov 14 1980)
Henri Frei and *La grammaire des fautes* (1929)

A German Swiss (from Baar near Zug) who wrote in French, Frei studied under Charles Bally, himself the pupil of Ferdinand de Saussure. His interest in the relationship between *langue* and *parole* is unsurprising, and indeed most of his articles appeared in the *Cahiers Ferdinand de Saussure*.

Frei’s most famous work, *La grammaire des fautes*, was based on the analysis of a written corpus, mostly built up of letters written by French soldiers during WWI. He examined ‘linguistic errors’ in this corpus and developed an account of the newer trends in the structure of informal French.
What the material in *Le livre*.... comprises

The material, collected though it was in pre-electronic times (a “steam corpus”), is semantically very wide-ranging and syntactically rich, and furnishes a conspectus of major structural frameworks in French.

2000 discrete sentences, each arranged in a subsection and preceded therein by a headword; c. 17000 words in the sentences.

Some of the sentences are Topic-Comment or Question-Answer pairs; others are single utterances.

Furthermore....

Provides numerous examples of the major phrase- and sentence types and some less frequent ones also (we find some instances of disjuncts, for instance: Bennett 1973 was especially concerned with this).

Transitive and ditransitive verbs are routinely provided with a series of actants rather than simply being listed as lexical items, phraseology is well exemplified.

The material is especially valuable in providing data on such matters as the prepositions which certain verb require.
Parallel texts as (proto)corpora

Considerable interest in this issue, as some texts (most famously Biblical texts, but also others) are available in multiple translations. **STUF** 60: 1 (2007) contained several treatments of this, including Stolz (2007) and Cysouw and Wälchli (2007).

This kind of approach has mostly been carried out on continuous texts, however, rather than on work like the “Frei corpus”.
Translations of this protocorpus.

• Chinese 1982 (*Fayu liang qian chang yongju*)
• Japanese 1971 (*Nihongo nisenbun*): Waseda University Press.
• -These were published versions of the translated sentences which Frei himself had collected.
• Unpublished versions of the sentences in Low Alemannic Baselditsch which Frei collected in 1952 are also known.
• Other versions of the sentences involved translations collected by other scholars. Translations were made into Latvian and Tajik and several other languages.
Test case: DONNER (to give)

• Regular 1\textsuperscript{st} conjugation verb
• #46 in frequency list for French
  https://duolinguist.wordpress.com/2015/01/06/top-5000-words-in-french-wordlist/
• Widely used in forming idioms
• 20 instances in this protocorpus
DONNER sentences in the Frei corpus (1966)

7  Ça lui a donné une telle frayeur qu’elle est devenue toute pâle.
    That gave her such a fright that she turned quite pale

42  Donne-moi quelque-chose à boire.
    Give me something to drink

140 (à lui) Quel âge lui donnez-vous?
    (to him) What age would you put him at?

225  Donne-moi de l’eau.
    Give me some water

326 (Nappe). Il faut la donner au blanchisseur.
    (cloth). You have yto give it to the launderer.

328 (Gilet) C’est mieux de le donner au teinturier.
    It’s best to give it to the dry cleaner,
Continued...

446  (Lui et moi). On s’est donné rendez-vous au débarcadère.
    (He and I). We met up at the jetty

555  (A propos d’un ouvrier mécontent). Je lui ai pourtant donné 50 francs, ça me semble plutôt généreux!
    (about an unhappy worker.) I’ve given him 50 francs, though – that seems rather generous to me.

564  Combien avez-vous donné pour cette maison?
    How much did you give for this house?

785  Ça m’a donné une frousse de diable.
    That scared me out of my wits.

891  Il a l’esprit très clair, aussi: ça peut donner un grand savant.
    He’s a very straightforward personality too; that can make for a great thinker.
Encore des phrases...

1071 *(A cet enfant). Eh bien, si on lui donnait un livre d’images?*
(to that child). Well, what if we gave her a picture book?

1074 *Quelle pièce est-ce qu’on donne maintenant?*
What play are they performing now?

1159 *(A la doméstique) Donne-moi du savon.*
(To the servant.) Give me some soap.

1180 *Cet arbre ne donne pas de fruits.*
This tree doesn’t bear fruit.
And finally....

1507 Votre maison donne sur quel côté?
What side does your house face?

1508 La maison donne sur la mer.
The house faces seawards.

1744 Donnez-moi un exemple pour montrer ce que vous entendez.
Give me an example to show what you are hearing.

1983 Si je vous donne ce modèle, vous pouvez l’imiter?
If I give you a model, can you imitate it?

1998 (Encre). Si tu m’en as, donne m’en un peu.
(Ink.) If you’ve got some, give me a bit.
Uses of DONNER

Tenses and other verb forms:
Present: 140, 1074, 1180, 1507, 1508, 1983
Imperfect: 1071
Perfect: 7, 555, 564, 785
Imperative: 42, 225, 1159, 1744
Infinitive: 326, 328, 891, 1998
Passive perfect: 446
Idiomatic or connotative uses: 1507, 1508 (connotative), 446 (part of an idiom).

Other uses in this collection are denotative.

All instances of DONNER show the use of both direct and indirect objects except sentences: 891, 1074, 1180, 1507, 1508.
Some translations of the corpus were published ‘as is’ (Bennett 1978)
Use of the Frei corpus in collecting endangered (Zakho Northeastern Neo-Aramaic) language data (Avinery 1974).
Frei corpus as basis of fuller description of the structure of a language (Boakye 1980)
Analysis of Frei corpus material to extract maximum lexical and morphological data (Galand and Zaouch 2011)
Pages from a vigorously-interrogated steam corpus: an excerpt from a morpheme list derived from the Frei corpus (Galand and Zaouch 2011)
Use of Frei corpus to document underdescribed variety of a well-known language: KiNgwana Swahili (Kaji 1985)
Frei corpus in use to document two dialects of the same language for contrastive purposes (Japanese)
Conclusion

Parallel protocorpora show their uses in the investigation of a wide range of structural and typological phenomena, bridging typology, philology and corpus linguistics.

Frei corpora may have less to say about collocations than comparable corpora, but they provide the materials for the construction of an extensive structural and lexical account of a language.
THANK YOU!
REFERENCES


