Applications of corpus-based research on language pedagogy:
Teaching personal opinion adverbials

Fryní Kakoyianní Doa & Elení Tziafa

University of Cyprus
frynidoa@ucy.ac.cy, tziata.eleni@ucy.ac.cy

Abstract
In this paper, we discuss the potential of a corpus-based approach in teaching. Moreover, we give an overview of the searchable online French-Greek parallel corpus (SOURCe), which aims to serve the needs of students of French as a foreign language and also to facilitate future linguistic research. This project is led by Fryni Kakoyianni-Doa and is fully funded by the University of Cyprus. The goal of the project is to provide online content on general education subjects to students and training in a cost-effective manner, as an Open Educational Resource. We demonstrate how these two corpora can be used for theoretical and pedagogical purposes by focusing on personal opinion adverbials.

Keywords: Source corpus, corpus-based research, open educational resource, French language acquisition, personal opinion adverbials

1. Introduction
In this paper, we discuss the potential of a corpus-based approach in teaching. As a case study, we present the characteristics of personal opinion adverbials, as attested in the Source Corpus, and their teaching perspectives.

Corpus-based studies allow a variety of factors to be investigated, such as frequency, semantic categories, grammatical structures, clausal positions, and individual items. Importantly, by studying a number of factors together, we can see how they are interrelated. When conducting analyses by hand or using a concordancer and keeping counts by hand, it is difficult to keep track of the interaction of numerous characteristics. Teachers need to know not only structural information about features that they are teaching, but also how native speakers use these structures under various conditions, if they want to promote communicative competence in their students. This combination of structural description and information about use is precisely what a corpus-based study can provide (Conrad 1999).
New tools are critical to creating a dynamic and engaging learning environment. Furthermore, according to Gabrielatos (2005:1), “changes in knowledge, skills and attitudes […] are needed for learners and teachers to take advantage of the opportunities offered by the availability of corpus resources”. Corpora are undoubtedly useful in language learning as they provide “a resource for vocabulary building in which the word is treated as part of a phrase rather than in isolation” (Hunston 2002: 106).

Our aim is to explore the potential uses of the SOURCe Corpus in language pedagogy. We will report on working with personal opinion adverbials, a subset of utterance-level French and Greek adverbials which express the personal opinion of the speaker towards the information transmitted, e.g. à mon avis [κατά τη (γνώμη + άποψη) μου] ‘in my opinion’, selon moi (κατ’ εμέ) ‘for me’, etc. It is examined, in particular, whether their morphosyntactic and semantic properties as well as the similarities and differences proposed for these adverbials in a previous work (cf. for French, Molinier 1984; Molinier & Lévrier 2000, Gross 1990, and for Greek, cf. Voyatzi 2006, Kakoyianni-Doa 2008; Kakoyianni-Doa, Voyatzi & Tziafa 2012), are attested in a real corpus. We will investigate their multiple characteristics so that we can gain a better understanding of their use. Specifically, for personal opinion adverbials we examine frequency, semantic category, grammatical structure, placement within the clause, the specific item used, and variation across registers (fiction and non-fiction discourse) as well as the interactions of these characteristics. Finally, we will outline its future perspectives and applications, in classroom materials and activities, discussing how it can be incorporated into effective learning resources.

2. Why did we choose to use corpora in the classroom?
Language changes and grows, and therefore there is a need for ‘currency’ and ‘up-to-dateness’ in the learning material as well. (McEnery & Wilson 1996; Leech 1992) “Educators must prepare for a technology-rich future and keep up with change by adopting effective strategies that infuse lessons with appropriate technologies.” (Shariat Madari, Mahdi & Mahdi, 2007: 278). The present-day interest in corpora has been described as a ‘corpus revolution’ (Leech 2000). As Mishan (2004) describes it, the key feature of the methodology is this ‘aura’ of research, and this research agenda is what gives the sense of authenticity.
Parallel and comparable corpora introduce translation in language learning, and they promote autonomy and language awareness. Using English-French data, Barlow (2000) sketches out how parallel corpora might be used in the classroom to sensitise students to the way in which different languages encode equivalent meanings.

It seems now that the general attitude towards translation has begun to change: those who discuss it in their studies argue that translation is a legitimate pedagogical tool in language learning, and claim that it deserves to be rehabilitated (Widdowson 1978; Harmer 1991; Ellis 1992; Bowen & Marks 1994; Ur 1996. Almost all authors seem to be in agreement that translation is most useful as a quick and easy way to present the meaning of words and contextualized items, and when it is necessary to draw attention to certain differences that would otherwise go unnoticed (Harmer 1991; Ellis 1992; Nunan & Lamb 1996).

3. The key-aspects of the Source Corpus

The Source Corpus is an online data base which aims to provide language teachers and learners in Cyprus and internationally with parallel corpora for use in the classroom, in order to promote computer assisted and data-driven language learning.

The key-aspects of Source Corpus include the following:

- It brings closer a well resourced language with a less resourced one.
- It can bring corpora in the classroom and establish a Data-Driven Learning (DDL) approach (Johns 1991).
- It promotes foreign language learning.
- It promotes the use of parallel corpora and translation.
- It is an Open Educational Resource.

This paper gives an overview of the searchable online French-Greek parallel corpus (SOURCe), which includes original texts with authentic translations (French-Greek and Greek-French), and overall, fictional and non-fictional texts. It is compiled for use in applied and theoretical linguistic research, but mostly aims to serve the needs of students of French as a foreign language and also to facilitate future linguistic research.

All translated texts are displayed in full sentences. Queries are based not only on words, but also on linguistic annotation. Most parallel corpora are not register-diversified; nevertheless, our objective is to include at least five different registers (Biber 1993), so that teachers and students may compare the results and the use of
each word or phrase in different contexts (e.g., literature, scientific, official, technical language).

These resources aim at improving learning and teaching (including 21st century skills, such as research conducted in electronic texts and use of language technology) for students and instructors, and are planned to be equipped with intuitive interfaces for visualizing and interacting with the data in order to ease their integration into the practice of teaching and learning. As Braun (2005, 2007) suggests, audiovisual materials for corpora of spoken language are going to be included, in order to give an idea of the overall communicative situation in which the materials were produced, and the corpora will be enriched with support materials, such as corpus-based learning activities and exercises.

Therefore, the resources that are going to be provided include:

- The corpora (OPUS, Gutenberg, digitised books)
- An online concordance as Key Word in Context (KWIC)
- Text analysis software
- Wordlists
- Collocation lists
- Exercises
- Links to useful resources
- Instructions
- A discussion about copyright issues and the public domain

4. Personal opinion adverbials

Our case study, using the corpus, are personal opinion adverbials, which present mainly homogeneous structures in French and Greek language, as long as they contain a 1st person deictic morpheme, as shown in previous research (Kakoyianni-Doa, 2008). In particular, these are multiword expressions containing nouns such as *avis* (άποψη + γνώμη), point de vue (σκοπία), jugement (κρίση) and possessive pronouns such as mon [δικ(ός-ή-ό) μου, μου], notre [δικ(ός-ή-ό) μας, μας], or personal pronouns such as moi (εμένα, μενα, εμέ), nous (εμάς, μας). Semantically, according to Molinier and Lévrier (2000), these adverbials indicate that the speaker presents his/her own subjective opinion, a personal statement, information with herself/himself as a source, and it may not be the absolute truth. For example:
Applications of corpus-based research on language pedagogy

(1a) À mon avis, il ne fera pas beau demain (mais je peux me tromper) (FR)

(1b) Κατά τη γνώμη μου, δε θα έχει οραίο καιρό αύριο (αλλά μπορεί και να κάνω λάθος) (GR)

(1c) In my opinion, weather won’t be nice tomorrow (but I may be wrong) (EN)

The Source Corpus is comprised of 5 registers: educational texts (“EDUC”), literature (“LIT”), political discourse texts (“POL”), scientific texts (“SCIENCE”) and technical texts (“TECH”). Using a sample corpus of one million words, divided in equal parts, we extracted the frequency of personal opinion adverbials across registers, as shown in the following diagrams:

Diagram 1. Frequency of personal opinion adverbials across registers

After disambiguation, it became even more obvious that the highest frequency is in political discourse, as shown in diagram 2:

Diagram 2. Frequency of personal opinion adverbials across registers, after disambiguation
The variation of personal opinion adverbials is also higher in the political discourse, as shown in Diagram 3:

![Diagram 3. Variation of personal opinion adverbials across registers](image)

There is a certain variety of personal opinion adverbials in political discourse, as shown in diagrams 4 and 5, for French and Greek language respectively. This is expected, because political discourse uses several persuasion techniques, where personal opinion adverbials play a prominent role. It is also a transcribed oral discourse, where the use of personal opinion adverbials is significantly frequent. The lower frequency in literature is also typical, since we do not expect to see clearly expressed the personal opinion of the author, unless there is a dialogue, or an imitation of oral discourse.

![Diagram 4. Frequency of personal opinion adverbials in political discourse (French Corpus)](image)
Diagram 5. Frequency of personal opinion adverbials in political discourse (Greek Corpus)

We observed that the personal opinion adverbials are much more frequent in Greek rather than in French; also, the adverbials in each language are not equivalent (as exact translations). For example, in French it is the adverbial *pour moi* (= for me, as far as I am concerned) with a frequency of 2,369, while in Greek the highest frequency (4,846) lays with *κατά τη γνώμη μου* (= in my opinion). The same goes with the other adverbials attested in the corpus. We see that, despite the common morphosyntactic features of these adverbials, as shown below in Table 1, the literal translation, word for word, is out of context, and cannot represent the linguistic reality or language use.

<table>
<thead>
<tr>
<th>Internal Lexical Structure</th>
<th>FRENCH</th>
<th>GREEK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preposition Personal Pronoun</td>
<td>4.329</td>
<td>1.539</td>
</tr>
<tr>
<td>Preposition Possessive Adjective Noun</td>
<td>3.810</td>
<td>8.346</td>
</tr>
<tr>
<td>Preposition Possessive Adjective Adjective Noun</td>
<td>33</td>
<td>111</td>
</tr>
<tr>
<td>Preposition Phrase</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Adverbial personnellement/προσωπικά</td>
<td>782</td>
<td>3.496</td>
</tr>
<tr>
<td>New structures (serendipitous discoveries in the corpus): Preposition Noun Adjective (FR)</td>
<td>344</td>
<td></td>
</tr>
<tr>
<td>Preposition Personal Pronoun Adverbial (GR)</td>
<td></td>
<td>117</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>9.310</strong></td>
<td><strong>13.611</strong></td>
</tr>
</tbody>
</table>

Table 1. Morphosyntactic structures of personal opinion adverbials

By searching the corpus, it was possible to confirm common morphosyntactic and semantic properties for these adverbials in Greek and French, as proposed in previous
work (Kakoyianni-Doa et al. 2012). First of all, scope (portée), which is external, as to the content of the phrase, and it depends on their lexical semantics.

(1) Για μένα υπήρξε μάλλον απογοηηεηική (GR).

(2) Elle a été plutôt frustrante pour moi (FR). (EN: In my opinion, she was rather disappointing).

(3) Κατά τη γνώμη μου, αυτή η απαίτηση δεν μπορεί να ικανοποιηθεί (GR).

(4) À mon avis, cette exigence est absurde (FR). (EN: In my opinion, this demand is absurd).

As regards syntax, the personal opinion adverbials are characterized by flexibility, a free placement in the sentence, at the head of the sentence, in the middle, or the end.

(5) À mon sens, il faut se concentrer sur trois domaines (FR).

(6) Κατά την άποψή μου πρέπει να σταθούμε σε τρεις τομείς (GR) (EN: In my opinion, we should focus on three fields).

(7) Formation qui, d’après moi, doit réunir trois conditions, trois aspects (FR).

(8) Κατάρτιση που, κατά τη δική μου άποψη, πρέπει να συμπεριλαμβάνει τρεις προϋποθέσεις, τρεις πτυχές (GR). (EN: A training that, in my opinion, should include three conditions, three aspects).

(9) Il a bien fait, selon moi (FR).

(10) Έπραξε καλός, κατά τη γνώμη μου (GR). (EN: He did well, in my opinion).

The personal opinion adverbial can be easily replaced by other adverbials or be removed without affecting the meaning of the sentence. The same goes for Greek, για μένα, which refers to a person itself cannot be replaced at all.

Finally, we observed that the use of comma before or after an adverbial (for example, in the case of για μένα) is not systematic in the corpus, in both languages.

(11) Για μένα ο αθεϊσμός είναι και η δύο (GR). (EN: For me atheism is both).

(12) Αυτό, για μένα, ήταν μια πολύ αφελής παρατήρηση, κατέληξε σε παρεξήγηση (GR). (EN: This, for me, was a very naive observation, ending in a misunderstanding).

(13) Pourtant c’est à mon sens le point le plus important (FR). (EN: In my opinion, this was the most important part).
Taking into consideration that according to the Common European Framework of Reference for Languages (CEFR, 2011), teachers are generally called upon to make decisions about classroom activities, which they can adjust in the light of learners needs. “They are expected to monitor the progress of pupils/students and find ways of recognising, analysing and overcoming their learning problems, as well as developing their individual learning abilities” (CEFR, 2011: 141).

According to CEFR (2011), learners are expected or required to learn from written texts

a) by simple exposure;

b) by simple exposure, but ensuring that new material is intelligible by inferencing from verbal context, visual support, etc.;

c) by exposure, with comprehension monitored and ensured by L2 question and answer, multiple choice, picture matching, etc. (CEFR, 2011)

Moreover, in accordance to CEFR (2011), learners are expected to learn by reading untreated authentic texts that the learner encounters in the course of direct experience of the language in use (daily newspapers, magazines, broadcasts, etc.) and authentic texts selected, graded and/or edited so as to be judged appropriate to the learner’s experience, interests and characteristics.

Therefore, there is need for ways to help teachers deal with these texts to construct exercises in a corresponding context, aiming to language acquisition based in communicative situations. Moreover, according to Hubbard (2012: 16) “using computer-based corpora and a concordancing program, the student is placed in the role of grammatical and lexical explorer rather than as a vessel for received knowledge”.

The following part illustrates how corpora and concordances can help learners acquire personal opinion adverbials.

5. Exercises based on corpus and concordances

Exercises are an important, vital and very practical aspect of using a corpus in classroom. Teachers need to know not only structural information about features that
they are teaching, but also how native speakers use these structures under various conditions, if they want to promote communicative competence in their students. This combination of structural description and information about use is precisely what a corpus-based study can provide.

New technologies, coupled with flexible and innovative teaching methodologies and didactics, offer very motivating ways of learning through translation exercises (in the widest sense), thus overcoming monotonous and one-dimensional qualities usually associated with translation (Witte, Harden & Ramos de Oliveira Harden 2009).

The learners, by means of the corpus, may recover the morphosyntactic and semantic function of adverbials in their contexts. They may learn to form the adverbials and understand the scope and the meaning. Parallel texts have the ability to place the learner in a new learning situation, a communicative one. Moreover, learners are encouraged to acquire the strategies of interaction and self-reflexivity. Corpus linguistics may contribute to more creative language learning, not passive any more, leading to a discovery and induction of syntactic rules (in accordance with the 5th competence of CEFR, the acquisition of methodology).

According to Crinon, Legros & Marin (2002), the learners should have at their disposal a large number of texts, relevant to those that is required for them to produce, that is belonging to the same genre or register. The register-diversified corpora correspond to this double demand. This also leads to a new attitude towards the consultation of texts, that of a learner-researcher of information.

We provide below some examples of exercises, in order to get the best out of the corpus and its concordances. There are different types of exercises (observation, comprehension (exercises 1, 2, 3, 4) and production (exercises 5 and 6).

**Exercise 1:**
Level: A1-A2 (CEFR)
Objective: Identification of personal opinion adverbials.
Methodology: Communicative/ task-based (CEFR)
Phase: Comprehension
Exercise for the disambiguation of the form για μένα / pour moi, as a personal opinion adverbial and referential για μένα / pour moi that refers to the person.
Observe the aligned sentences. Beginning by replacing για μένα with another adverbial (κατά τη γνώμη μου, κατά τη δική μου άποψη, etc.) in the sentences in your native language, identify by which για μένα is expressed:

– Opinion, as a result of the speaker thinking
– a reference to a person

Then match Greek and French sentences and do the same exercise for French sentences (replace pour moi with another adverbial (à mon avis, d’après moi, etc.) in the sentences).

<table>
<thead>
<tr>
<th>Για μένα αυτό είναι ο λάθος δρόμος.</th>
<th>Ne pleure pas pour moi.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Μην κλαίει για μένα.</td>
<td>Tu travailles pour moi, à présent.</td>
</tr>
<tr>
<td>Τώρα δουλεύεις για μένα.</td>
<td>Pour moi, la méthode est mauvaise.</td>
</tr>
</tbody>
</table>

Exercise 2:
Level: B1 (CEFR)
Objective: Identification of personal opinion adverbials in full authentic texts.
Methodology: Communicative / task-based (CEFR)
Phase: Comprehension

Exercise for the disambiguation of the form pour moi, as a personal opinion adverbial and referential pour moi that refers to the person.

Observe the concordances and identify the forms pour moi that may refer 1) to the person as a whole and 2) the personal opinion of the speaker.
The concordances from the Greek part of the corpus are also available (see below), so that students may better understand the different uses of the adverbial. We have the same ambiguity. “Reception strategies involve identifying the context relevant to it, activating in the process what are thought to be appropriate schemata. During the process of receptive activity cues identified in the total context and the expectations in relation to that context set up by the relevant schemata are used to build up a representation of the meaning expressed and a hypothesis as to the communicative intention behind it” (CEFR, 2011: 72).
Exercise 3:
Level: A2
Objective: Identification of personal opinion adverbials
Methodology: Communicative / task-based (CEFR)
Phase: Comprehension
Exercise for matching personal opinion adverbial equivalents in the two languages.
Match the following French and Greek phrases in order to reconstitute the original parallel corpus.

<table>
<thead>
<tr>
<th>1. Κατά τη γνώμη μου, και τα τρία αυτά στοιχεία με &quot;ε&quot; εγγυώνται αυτή την αειφορία.</th>
<th>a. À mon avis, il faut que, dans tous les États membres, la limite du taux soit ramenée</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Θεωρώ σημαντικό να θεσπιστεί από όλα τα κράτη μέλη ο περιορισμός του ανώτατου ορίου</td>
<td>b. À mon avis, ces &quot;trois E&quot; garantissent cette durabilité</td>
</tr>
<tr>
<td>3. Ο πυρήνας της έκθεσης, κατά τη γνώμη μου, είναι ότι εντάσσει αυτή την υπόδειξη της εφαρμογής της αρχής &quot;ο ρυπαίνον πληρώνει&quot; στην ειρύτερη αρχή της πληρωμής από τον χρήστη.</td>
<td>c. La résolution condamnatoire d'aujourd'hui - exagérément indulgente à mes yeux - constitue toutefois une réaction positive, évidente</td>
</tr>
<tr>
<td>4. Το σημερινό καταδικαστικό ψήφισμα - υπερβολικά επιεικές κατά τη γνώμη μου - αποτελεί, εν τούτοις, αυτονόητη και θετική αντιδραση</td>
<td>d. À mon avis, le point cardinal du rapport fait en sorte que cette indication d'application du principe du pollueur-payeur vienne s'insérer dans le concept plus large de l'utilisateur-payeur.</td>
</tr>
</tbody>
</table>

Advantages of matching sentences is that they get at a student’s ability to identify relationships and classify things, this means we can get not only at recognizing targets but also understanding targets (such as, exemplifying, classifying, and comparing).

Exercise 4:
Level: B1
Objective: Identification of personal opinion adverbials and their syntactic position
Methodology: Communicative / task-based (CEFR)
Phase: Comprehension
Exercise for the comprehension of particularities of use in each language.
Observe the parallel corpus and
a) Indicate the corresponding uses of personal opinion adverbials in French and Greek.

b) Indicate the syntactic position of the adverbial in each language.

c) Indicate the parts of speech that precede or follow.

<table>
<thead>
<tr>
<th>English (Greek)</th>
<th>French (French)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Katá tì gnóììì înì μìnì, και τά τρία αυτά στοιχεία με &quot;e&quot; εγγυώνται αυτή την αειφορία.</td>
<td>À mon avis, ces &quot;trois E&quot; garantissent cette durabilité.</td>
</tr>
<tr>
<td>Θεωρώ σημαντικό να θεσπιστεί από όλα τα κράτη μέλη ο περιορισμός του ανώτατου ορίου</td>
<td>À mon avis, il faut que, dans tous les États membres, la limite du taux soit ramenée</td>
</tr>
<tr>
<td>Ο πυρήνας της έκθεσης, κατά τη γνώμη μìnì, είναι ότι εντάσσει αυτήν την υπόδειξη της εφαρμογής της αρχής &quot;ο ρυπαίνον πληρώνει&quot; στην ευρύτερη αρχή της πληρωμής από τον χρήστη.</td>
<td>À mon avis, le point cardinal du rapport fait en sorte que cette indication d'application du principe du pollueur-payeur vienne s'insérer dans le concept plus large de l'utilisateur-payeur.</td>
</tr>
<tr>
<td>Το σημερινό καταδικαστικό ψήφισμα - υπερβολικά επικείμενο κατά τη γνώμη μìnì - αποτελεί, εν τούτοις, αυτονόητη και θετική αντίδραση</td>
<td>La résolution condamnatoire d'aujourd'hui - exagérément indulgente à mes yeux - constitue toutefois une réaction positive, évidente</td>
</tr>
</tbody>
</table>

In this exercise, learners can find out in what order the elements have been produced.

The reader’s eye is free to move over the text in any way, possibly following the linear sequence in strict order, as a child learning to read will generally do. Skilled, mature readers are much more likely to scan a text for highly information-bearing elements in order to establish an overall structure of meaning and then return to read more closely – and if need be to re-read a number of times – such words, phrases, sentences and paragraphs (CEFR, 2011: 97).

In previous exercises the learners had to understand functions, ambiguities, syntactic positions, differences and similarities between Greek and French language. In exercise 5 the learner has to reconstitute phrases of the corpus that contain the personal opinion adverbial which means that he should place the adverbial in its proper position, and so the other parts of speech. According to CEFR (2011), the B1 learner “can use a variety of strategies to achieve comprehension, including listening
for main points, checking comprehension by using contextual clues” (CEFR, 2011: 81).

**Exercise 5:**
Level: B1
Objective: Re-constitute phrases of corpus that contain the personal opinion
Methodology: Communicative / task-based (CEFR)
Phase: Production
Exercise for phrase delimitation, ordering the phrase’s elements, placing the adverbial properly.
Complete and re-constitute the phrases of the following *corpus* that contain the personal opinion adverbial “à mon avis”.

Similarly, a written text may be carefully planned in exercise 6 according to models (from previous exercises, comprehension) so as to appear to be spontaneous, “yet to ensure that an essential message is effectively conveyed under the different conditions that constrain the reception of speech” (CEFR, 2011: 97).

**Exercise 6:**
Level: B2 of CEFR
Objective: Use of personal opinion adverbials
Methodology: Communicative/task-based (CEFR)
Phase: Guided expression
Model-based Production
You give a speech in the Council of Europe about a policy that could relieve the financial problems in South Europe. Most importantly, you have to state your
personal opinion. Use, as an aid, the Europarl Corpus. Search for personal opinion adverbials.

In exercise 6, the learner “can summarise and give his or her opinion about a short story, article, talk, discussion, interview, or documentary and answer further questions of detail” (CEFR, 2011: 81), and also

the learner can understand enough to manage simple, routine exchanges without undue effort. Can deal with practical everyday demands: finding out and passing on straightforward factual information. Can use an idea of the overall meaning of short texts and utterances on everyday topics of a concrete type to derive the probable meaning of unknown words from the context (CEFR, ibid).

6. Future perspectives and applications
Our main objective is to develop a knowledge base that will serve to assist teachers to find out about, adapt and apply existing tools. Moreover, the corpus page could be a virtual space for gathering and collating information on corpora, identifying software, tools, and also relevant publications and training information. Nevertheless, the corpus construction could be completed with an exercise generator, in order to assist teachers in the development of relevant exercises. Having in mind that a corpus is not the end point for teaching and learning but it is the beginning, we aim to make it the beginning of a creative source of ideas for French language teachers and learners.

References