TOWARDS A CORPUS PRAGMATICS OF ELF THROUGH SEMI-AUTOMATED ANNOTATION SYSTEMS

LAURA CENTONZE
UNIVERSITÀ DEL SALENTO

Abstract – The present paper illustrates an undergoing doctoral research project (Centonze, forthcoming) aimed at introducing a novel approach to the description of spoken discourse in ELF in migration settings which combines corpus linguistics, corpus pragmatics (Aijmer and Rühlemann 2015) – a relatively new research area in the field of language and discourse studies – with the most recent techniques of quantitative/qualitative analysis and corpus annotation by means of semi-automated software tools. More specifically, the project focuses on the pragmatic annotation of speech acts from an ELF perspective and on the analysis of speech acts in their frequencies and collocations in a study corpus by means of *DART* (the *Dialogue Annotation Research Tool* v. 1.1., Weisser 2015), i.e. a research tool which, among other things, includes the functions of both POS (Part-Of-Speech) tagging and pragmatic annotation of spoken discourse. The corpus which is being taken into consideration is an under-construction corpus which will be referred to as the *ELF MiDo Corpus* (English as a Lingua Franca in MIgration DOmains corpus) and consists of over 50,000 words of conversation between asylum seekers and intercultural mediators in symmetrical contexts of interaction. All the different corpus interviews and interactions are transcribed and annotated according to a basic .XML mark-up scheme which proved to be a necessary condition for the whole corpus to be properly scanned for analysis through the *DART* interface. The aim of the present research study is to assess – by illustrating two case studies taken from the corpus – the use of *DART* for the pragmatic description of discourse in ELF and to verify the extent to which (semi-)automated software tools like this can effectively capture pragmatic change in interactional settings.

Keywords: ELF; corpus pragmatics; *DART*; speech acts; pragmatic annotation.

*L*ingue e
*L*inguaggi

1. Introduction[1]

The use of *English as a Lingua Franca* (henceforth ELF; Seidlhofer 2001) on the part of speakers whose native language is other than English has been gaining momentum in the last decades, especially due to the migration flows of people from their home countries to Europe in order to get a better life and better job opportunities for themselves and their families. As a consequence, there has been an urgent need to train people to provide free-of-charge consultancy services and other related facilities to migrants and asylum seekers both worldwide and locally, and to provide adequate resources for the adoption of a shared variety of English which would act as a *lingua franca* among people belonging to diverse lingua-cultural backgrounds (Cogo *et al.* 2011). A high number of non-profit associations are thus emerging in a way to facilitate such processes and find the most suitable way to grant a permit to stay to migrants and asylum seekers, together with a range of additional services which include facilitating the search for a job and the successful integration of migrants within society, also thanks to specific training courses aimed at enhancing their knowledge of the culture and traditions relating to the hosting country.

By considering the above-mentioned socio-cultural and linguistic scenario, the aim of the present paper is to assess the feasibility of (semi-) automated methods adopted for the pragmatic analysis of spoken discourse, to apply such methodology to an under-construction corpus of interactions between asylum seekers and intercultural mediators in institutional encounters (the *ELF MiDo Corpus*, i.e. the E*nglish as a* L*ingua* F*ranca in* M*i*g*ration* D*omains Corpus*, Centonze forthcoming) and to make it available in its annotated version for the analysis of speech acts and other pragma-linguistic features such as turn-taking, syntactic categories of verbs and so forth. By adopting a corpus-pragmatic approach, we provide an integrated model for the analysis of such interactions, which combines the most recent techniques of corpus linguistics, corpus pragmatics as well as POS-tagging of digitalized discourse and which could be of help for the training of intercultural mediators and the identification of pragmatic patterns in ELF conversations in migration contexts. More specifically, by means of two distinct case studies, the present paper provides grounds for the necessity to improve current semi-automated software options available for the retrieval of the pragmatic function of speech acts and to point to their strengths as well as their weaknesses. In order to fulfill our aim, we have analyzed the tags that

---

[1] The research project was presented on occasion of the ELF symposium "English as a Lingua Franca: Expanding Scenarios and Growing Dilemmas" which took place at Sapienza University (Rome, 6-7 April 2017).

Lingue e Linguaggi

were associated to each relevant speech act within the study corpus sections by means of *DART* (the *Dialogue Annotation Research Tool v.1.1*, Weisser 2015) and focused on two case studies taken from it.

The following sections shall respectively deal with the theoretical background upon which the present study is based (Section 2); the description of the under-construction corpus which constitutes the object of the present study (Section 3); Section 4 shall provide a description of the *DART* software tool which was applied for the analysis of speech acts, together with its functionalities; Section 5 shall present the two case studies where *DART* was applied and then we shall draw conclusions relating to them and provide points for further research in the field.

2. Theoretical background

2.1. Speech act theory

Since the purpose of the present study is to assess the feasibility of (semi-) automated means for the retrieval of speech acts and, more specifically, the adoption of the *DART* software tool in this respect to fulfill this aim, it goes without saying that the theoretical background which was taken into consideration as a bedrock is – first of all – represented by Austin's (1962) and Searle's (1969) theory for speech acts. With reference to the present study, we shall consider both the concept of speech act in its broader sense and definition, together with the three dimensions that a speech act incorporates. Searle's explanation is emblematic and makes it clear what a speech act actually represents and how it becomes contextualized in conversational settings:

> The unit of linguistic communication is not, as has generally been supposed, the symbol, word or sentence, or even the token of the symbol, word or sentence, but rather the production or issuance of the symbol or word or sentence in the performance of the speech act. To take the token as a message is to take it as a produced or issued token. More precisely, the production or issuance of a sentence token under certain conditions is a speech act, and speech acts […] are the basic or minimal units of linguistic communication. (Searle 1969, p. 16)

Starting from what a speech act is *not*, what transpires from Searle's definition of speech act is the extent to which its notion is so concrete that its characteristics may be inferred from the relevant context in which it occurs (in Searle's words, 'the *production or issuance* of the symbol or word or sentence in the performance of the speech act […] the *production or issuance* of a sentence token *under certain conditions*', emphasis added). With regard

*L*ingue e *L*inguaggi

to this, Searle makes a distinction between three dimensions of speech acts which constitute three different levels of their realization: a *locutionary act*, which consists of the structure of a certain utterance, which incorporates an *illocutionary force*, residing in the communicative intent and objective of a given utterance, and a *perlocutionary effect*, which represents the effects of an utterance on the interlocutor. For the purposes of the present analysis we shall consider these three distinct phases of the speech act realization in order to assess whether the *DART* software tool applied to discourse in ELF is able to seize them and, if so, to what extent it is accurate.

2.2. Corpus pragmatics and its relevance to ELF

Corpus pragmatics is a relatively new discipline in the field of applied linguistics which is thriving over the last decades and combines the study of corpora – whether digitalized or not – and the analysis of pragmatics in specialized discourse. What makes it innovative as a discipline and is gradually making it emerge as a free-standing field of study is the corpus-assisted approach that characterizes it: as Aijmer and Rühlemann (2015, p. 3-9) suggest, this new trend in the analysis of discourse has brought together two sub-disciplines which are characterized by different methodologies: whilst – in Aijmer and Rühlemann's terminology (*ibidem*) – pragmatics keeps an 'horizontal-reading methodology' which is based upon the analysis of small texts that are easy to read and analyse, the methodology adopted in corpus linguistic studies is one of 'vertical reading', where K*ey* W*ords* I*n* C*ontext* (KWIC) are analyzed in a set of texts – usually very huge sets of data – in order to explore and identify the most occurring patterns.

Corpus pragmatics acquires much more relevance within the framework of the present paper, which considers speech acts in their three-dimensional function and, most of all, in their pragma-linguistic features in a corpus of conversational turns which are retrieved semi-automatically by means of *DART*.

3. The study corpus

The study corpus that is taken into consideration for the present paper consists of a collection of recorded oral interviews between asylum seekers and intercultural linguistic mediators carried out at the local *Consiglio Italiano per i Rifugiati* (*Italian Council for Refugees*) in Lecce as well in other centres in the province of Lecce (including Lecce and the municipality of Andrano, where there is a centre for migrants and asylum seekers in which they are included under certain specific conditions of emergence and under EU-funded projects) which give hospitality and psychological – as well as

administrative – support to migrants and asylum seekers in their quest for asylum and for their permit-to-stay renewal procedure and other migration-related issues (e.g. accommodation; job search; help with administrative formalities; filling in the form for the Italian for Foreigners test). Migrants and asylum seekers taking part into the interviews come from either Mali or Ghana, whereas intercultural linguistic mediators that were involved in the interviews had all been trained as part of a one-year post-graduate master programme in *Mediazione Linguistica Interculturale in Materia di Immigrazione e di Asilo* (*Intercultural Linguistic Mediation in Migration and Asylum-seeking Contexts*, our translation) at the Università del Salento (Lecce, Italy) and were all completing a work-experience module as part of their on-site training. The following table illustrates the breakdown of the corpus that is going to represent the primary set of data under analysis, which was labelled as the E*nglish as a* L*ingua* F*ranca in* Mi*gration* Do*mains* (henceforth *ELF MiDo*) corpus:

	No. words	Speaker's origin	Topic
1	2,803 words	Mali	Culture; job opportunities; migration
2	3,055 words	Ghana	Migration; permit to stay; family
3	2,841 words	Ghana	Family; leisure activities; money
4	3,989 words	Mali	Hardship of life; problems; migration
5	3,277 words	Mali	School; family reunification
6	2,456 words	Ghana	Home country; host country; culture
7	3,466 words	Ghana	Money; family; children
8	2,279 words	Mali	Everyday life; family; home country
9	4,765 words	Mali	Family; children; home country; reunification
10	3,971 words	Ghana	Traditions; home vs. host country
Tot.	**32,902 words**		

Table 1
Breakdown of the *ELF MiDo Corpus.*

As can be seen in the table provided above, the corpus consists of 10 interviews of approximately 35 up to 50 minutes in length and the topics which constitute the content of each interview are diversified and most of the times involve a report of the migrants' experience as they cross the Mediterranean and reach Italy – either in order to reach other countries (e.g. Germany) or to settle down and start a new life. More specifically, they generally report on key facts that are peculiar to their own experience in Italy together with some anecdotes concerning the cultural differences and problems they have had to face since their arrival in Italy – sometimes these narrations are curious, sometimes embarrassing, sometimes simply sad

vicissitudes. However, as can be seen, the corpus definitely does not constitute an extremely large set of data if compared to more ambitious projects such as the ELFA corpus (Mauranen *et al*. 2008) and the VOICE (Seidlhofer *et al*. 2013). Notwithstanding this, if we consider the specific aim of the present study which is a methodological exploration of annotation procedures by means of semi-automated software tools, this does not represent a disadvantage that prevents us from fulfilling this aim.

4. *DART* and its main functionalities

The *Dialogue Annotation and Research Tool* was developed by Martin Weisser at Guangdong University of Foreign Studies, with an aim to providing a useful tool for the automatic annotation of transcribed spoken interactions as well as for the post-editing of annotated data. The tool represents the offspring of two previous projects which aimed at providing some guidelines and resources for annotation, i.e. *The Expert Advisory Group on Language Engineering Standards* (EAGLES) *WP4 1997-1998* and the *Speech-Act Annotated Corpus of Dialogues* (SPAAC) *2001-2002*. The need for *DART* derived from the limitations of *SPAAC*, one of which was represented by its highly monolithic approach to data, where there was "no separation of linguistic intelligence and output display" (Weisser 2014). *DART* goes further by providing a model characterized by a "strict separation of processing and linguistic analysis routines" (Weisser 2014)[2] and by a more flexible approach which allows one to create new tags and thus personalise research methodologies. In the following figures, some insights into the *DART* interface are provided, together with its sections and uses.

---

[2] Weisser (2014) is a PowerPoint presentation. Both quotations were drawn from Slide 6 (*Design Background – 3*).
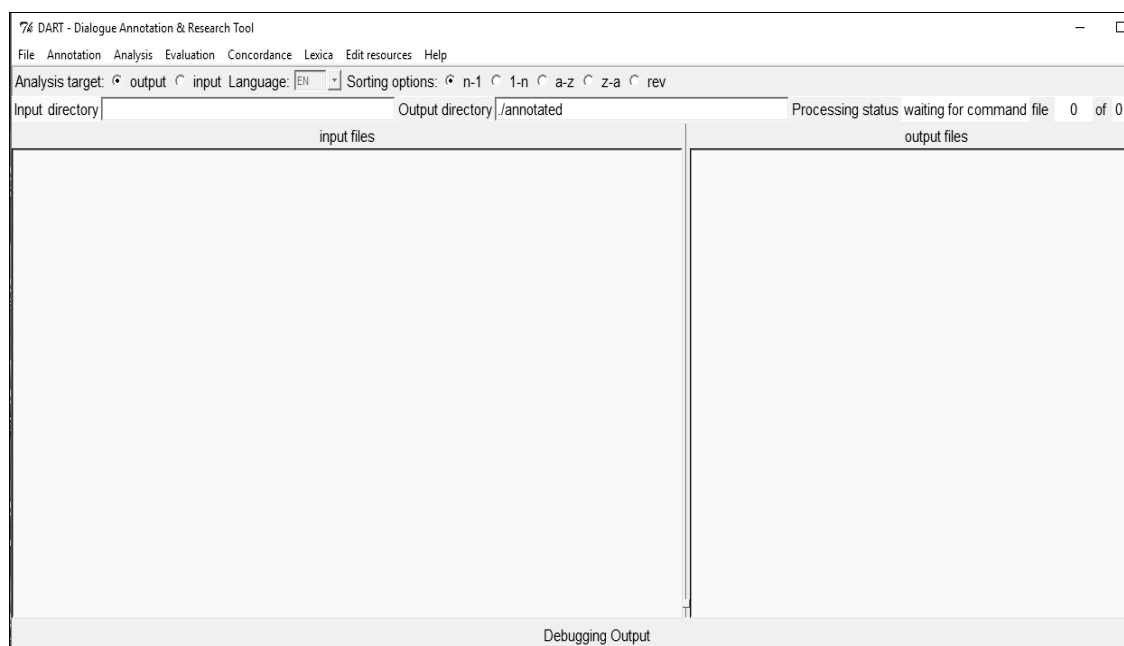
Figure 1
The *DART* interface.

As can be seen in Figure 1, the *DART* interface consists of a *Menu* with several options: from *File* one can upload both single .XML files as well as folders containing a series of files to be processed; the *Annotation* command allows for the annotation of files from two different perspectives: *POS* (Part Of Speech) tagging and *Pragmatic* (which implies the speech act tagging); the *Evaluation* command is the tool which allows us to carry out statistical analysis on speech acts and other parts of discourse, depending on whether we decide to carry out a POS analysis or a pragmatic one; the *Concordance* command identifies collocations for each item that is found in the relevant tagged corpus; the *Lexica* command allows us to see words by tag, whereas the *Edit Resources* command helps us take notes concerning the corpus itself. As one can see, the interface is divided up in two parts: a left one, i.e. *Input Files*, and a right one, i.e. *Output Files*. The *Input File* section represents the first step towards the analysis of corpora in *DART*: the felicity condition in order to carry out analysis in *DART* is the upload of files in .XML format; after being uploaded via the *File Menu*, such files can be then edited using the *Input Files* section. Once the file has been uploaded, a link to it is generated in the left section (i.e. *Input Files*), as Figure 2 shows:
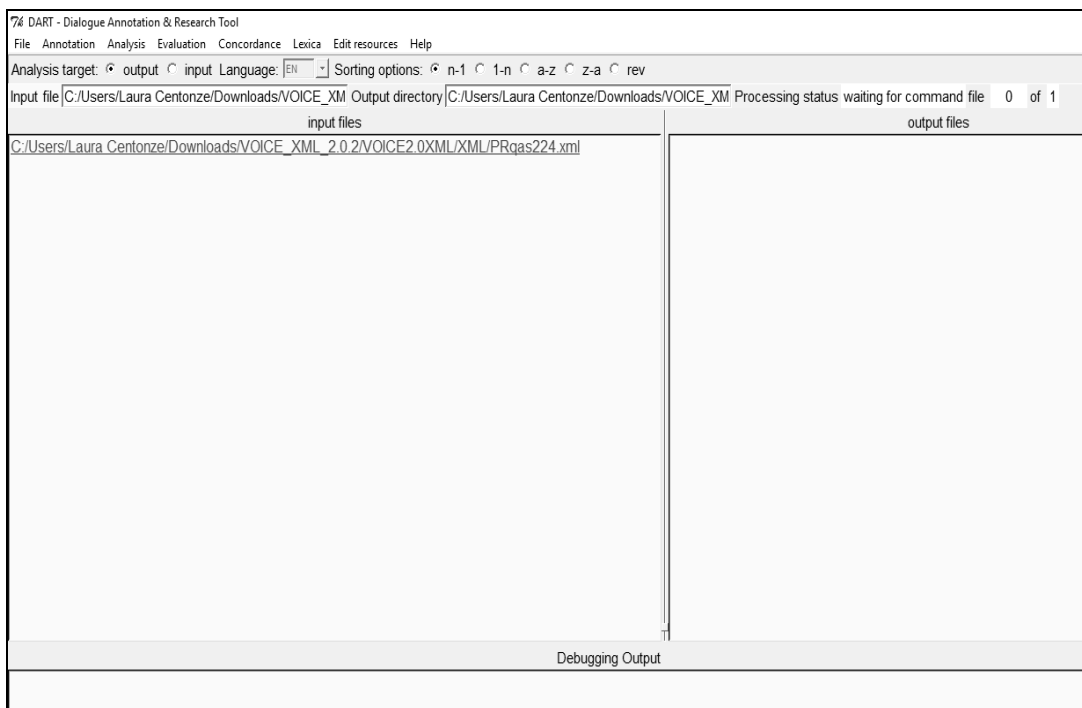
Figure 2
Uploading a file in the *Input Files* section.

After clicking on the link in the *Input File* section, a window like the one below opens (Figure 3); original files can then be edited and an .XML declaration (i.e. <?xml version="1.0?"> <dialogue corpus="name of corpus file" id="number of file" lang="en") can be added. This represents a necessary condition for the file to be processed properly.
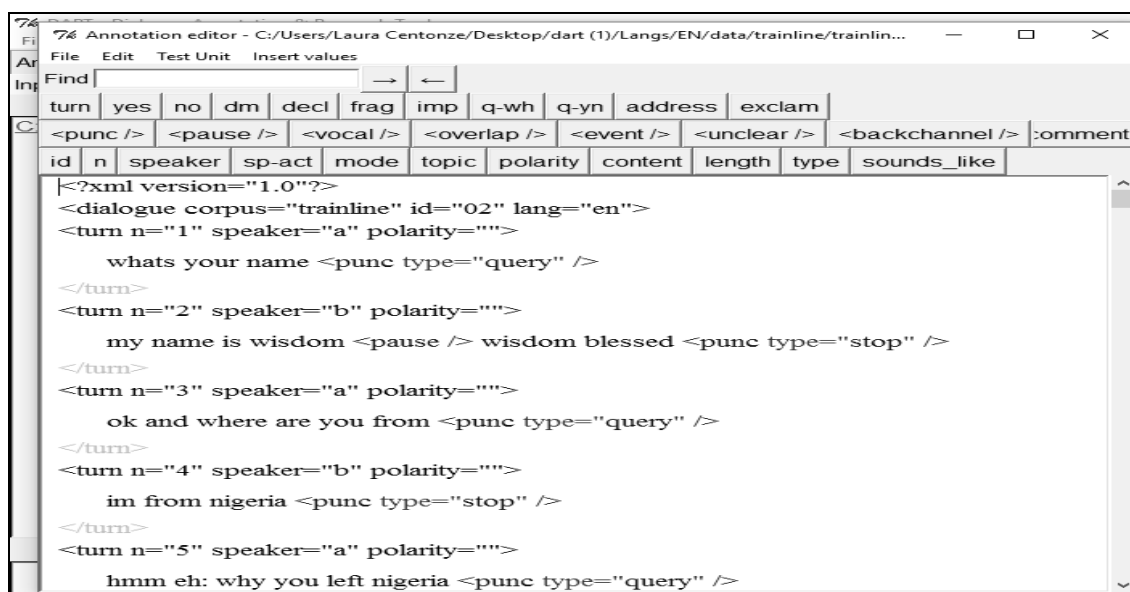

Figure 3
An example of preliminary processing of files in *DART*.

The one above only represents a sample and, as can be seen the dialogue is divided up into turns, which are numbered and each of them is separated by a punctuation mark which varies according to the function of each sentence (e.g. *question* and *statement* respectively "query" and "stop"). A full list of all tags can be found in the Appendix.

Once the whole file is divided up into turns, by means of the *Test Unit* command it is possible to verify the accuracy and conformity of each tag. After this preliminary action is carried out, we save the file and close the editing window; afterwards, we select *Annotation>Pragmatic* from the main menu and the following appears on the screen:
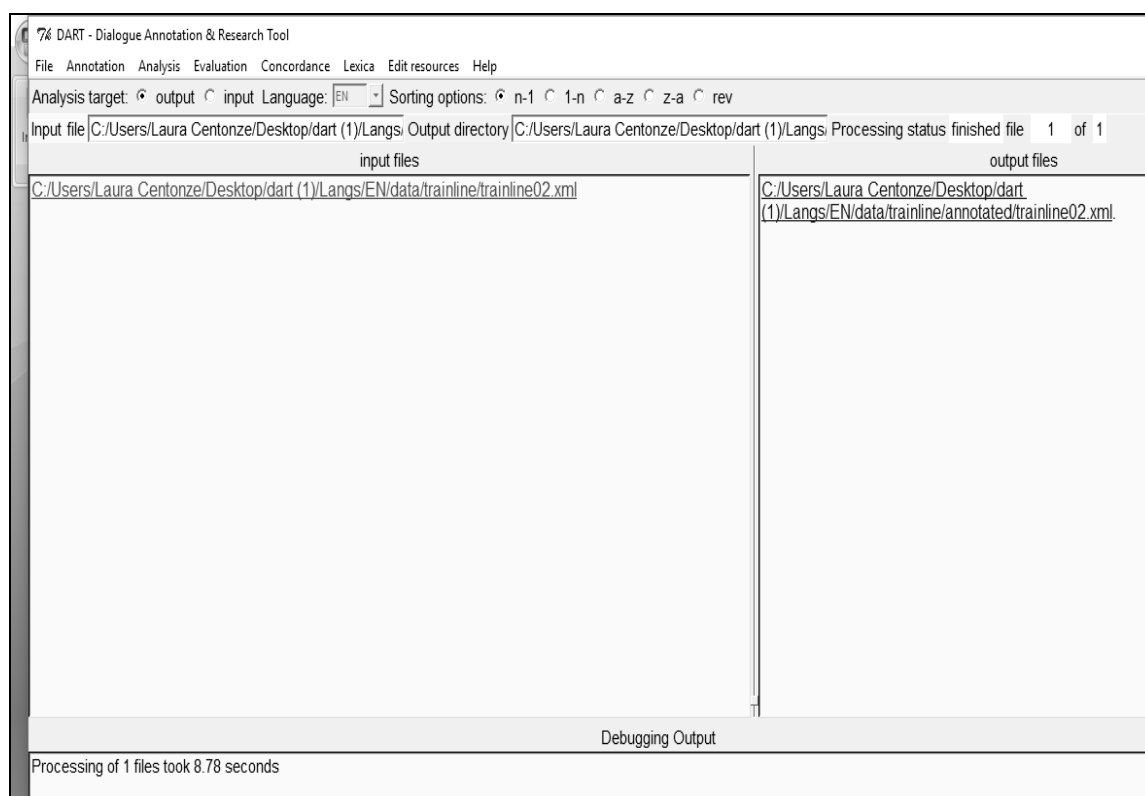

Figure 4
Pragmatic processing of files in *DART*.

Once the link provided on the right is opened, the file which has been processed and annotated pragmatically in *DART* can be displayed (Figure 5):
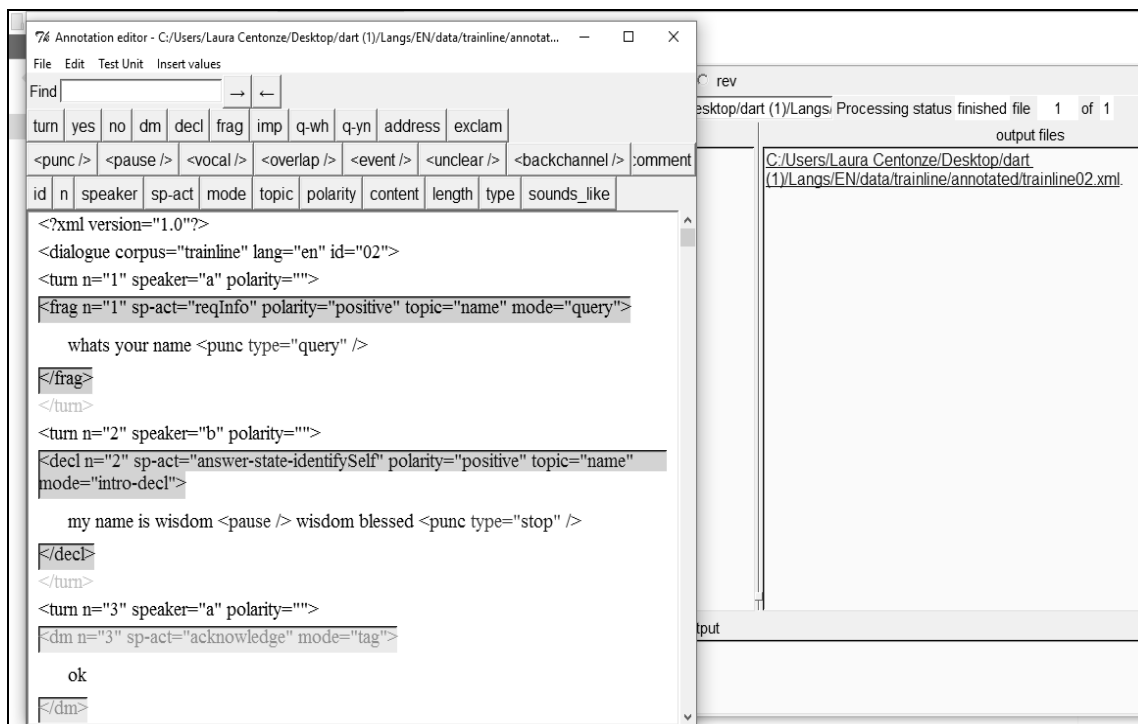
Figure 5
Example of pragmatically-annotated file in *DART*.

As can be seen above, once the processed file is opened the subdivision in turns can be displayed together with a preliminary identification and attribution of speech acts for each fragment. For instance, if the first two turns are taken into consideration, the outcome is the following:

```
<?xml version="1.0"?>
<dialogue corpus="mido" lang="en" id="02">
<turn n="1" speaker="a">
<frag n="1" sp-act="reqInfo" polarity="positive"
topic="name" mode="query">
whats your name <punc type="query" />
</frag>
</turn>
<turn n="2" speaker="b">
<decl n="2" sp-act="answer-state-identifySelf"
polarity="positive" topic="name" mode="intro-decl">
my name is $$$ <pause /> $$$ ### <punc type="stop" />
```

As can be observed, the speech act attributed by *DART* to the first turn corresponds to "reqInfo", i.e. a request for information on the part of the speaker, whereas the second turn contains an "answer-state-identifySelf" speech act. Moreover, the pragmatic annotation of the dialogue also contains some additional information, such as the type of sentence (whether it is a question/statement), polarity (positive/negative),

topic as well as mode. An exhaustive list of tags which can be attributed in *DART* is provided in the Appendix.

5. Testing *DART* for speech act identification and recognition

In order to assess the feasibility of *DART* as concerns the identification and recognition of speech acts in the ELF MiDo Corpus, we took a 3,000+ sample from the study corpus and ran the *DART* software tool in search for speech act frequencies in that specific section. Preliminary findings are reported below, which include speech act functions and frequencies >10:

Syntactic mode	Speech act function	Frequency >10
dm	acknowledge	74
frag	state	34
dm	exclaim	15
frag	reqInfo	14
decl	state	11
frag	**Unrecognized**	**45**

*`dm:` discourse markers; `frag:` fragments (e.g. ungrammatical sentences); `decl:` declaratives

Table 2
Speech act functions in a sample from the study corpus.

As can be seen from the speech act frequencies in the specific sample of the study corpus, there is a higher number of speech acts with the function of `acknowledge` (i.e. to confirm a status of things or some previous statement, 74 items found), with an overall prevalence of `dm` (discourse markers) over `frag` (fragments); if we have a look at fragments, we can see a high frequency of *unrecognised* speech acts, i.e. speech acts for which the *DART* software tool failed to retrieve a pragmatic function. This latter category has represented the focus of the following two case studies, which enabled us to point to some of the weaknesses of the program as regards the accuracy of speech act function retrieval. What is proposed here in order for the study corpus to be annotated accurately is a three-stage model, which implies (1) a preliminary automatic retrieval of speech act functions by means of *DART*, (2) an intermediate phase, which consists of reformulation techniques that are typical of a *retrospective verbal report* approach (Ericsson and Simon 1984) and which inevitably takes into consideration the *text* vs. *discourse* dichotomy highlighted in Widdowson (1996a), and (3) a third phase, during which the data has been predisposed for investigation. The second phase (i.e. retrospective verbal report) plays a pivotal role in the

process of re-definition of unrecognised tags and, in order to carry out this, ten intercultural linguistic mediators were asked to paraphrase strings of conversational turns which fell under the 'unrecognised' category according to *DART*, after being given up to 8 lines before and after the relevant speech act in order to be able to interpret each of them appropriately. The following two case studies illustrate three distinct examples where the 'unrecognised' speech act function was re-defined.

5.1. *Case study 1:* `sp-act"confirm"` *and* `sp-act"reqConfirm"`

The first instance that we considered in order to test the above mentioned model with special reference to the retrospective verbal report phase relates to the re-definition of unrecognized tags, i.e. those for which the *DART* software tool was unable to attribute a tag function. The example below is taken from a conversation between a migrant from Mali (b) and an intercultural mediator (a) which is aimed at gathering information concerning the period spent by the migrant at accommodation centers administered by non-profit organizations. The transcript was first reported in its 'unidentified' version for speech act function, then we applied the intermediate phase of retrospective verbal report by asking the ten intercultural mediators involved in the project to paraphrase and thus provide themselves the tag which was thought to be appropriate to the relevant context:

```
</turn>
<frag n="846" sp-act="" mode="decl">
rinascita si si si si <punc type="stop" />
</frag>
</turn>
<turn n="497" speaker="a">
<frag n="847" sp-act="" polarity="positive" mode="decl">
rinascita ah? <punc type="stop" />
</frag>
</turn>
<turn n="498" speaker="b">
<frag n="848" sp-act="stateReason" topic="time-spell"
mode="reason-decl">
when i leace de project because when de took us in eh
lampedusa mhm no in manduria dei took us to copertino <punc
type="stop" />
</frag>
</turn>
```

The two *unidentified/unrecognized* speech act functions are highlighted in grey and the intercultural mediators were given contextualized strings of turns which allowed them to reformulate in their own words the pragmatic

function associated to the speech act and then compare their answers against the speech act taxonomy provided by Weisser (2015) for *DART* v.1.1. and which can be found in the Appendix. The outcome is represented below:

```
</turn>
<frag n="846" sp-act="confirm" mode="decl">
rinascita si si si si <punc type="stop" />
</frag>
</turn>
<turn n="497" speaker="a">
<frag n="847" sp-act="reqConfirm" mode="decl">
rinascita ah? <punc type="stop" />
</frag>
</turn>
<turn n="498" speaker="b">
<frag n="848" sp-act="stateReason" topic="time-spell"
mode="reason-decl">
when i leace de project because when de took us in eh
lampedusa mhm no in manduria dei took us to copertino <punc
type="stop" />
</frag>
</turn>
```

The name *rinascita* refers to an organization that is available locally, helping migrants get accommodation and other related services. What the migrant (speaker b) is doing in `frag n=846` is to confirm what the intercultural mediator (speaker a) has elicited before that specific turn; probably the migrant had not been able to remember the name of the association and the intercultural mediator, who is aware of the local situation concerning services and facilities available to migrants, has made an attempt to help him/her by providing a series of names. The `sp-act="confirm"` is what the intercultural mediators provided as a final tag; likewise, the intercultural mediator is – in the following turn `frag n=847` – again asking for confirmation on whether s/he has understood the name properly. The tag which all intercultural mediators have agreed upon is "`reqConfirm`".

5.2. *Case study 2: `sp-act"reqInfo"`*

In the second case study, the following excerpt was taken from the study corpus which includes a conversational exchange between the migrant and the intercultural mediator, who is asking about the migrant's life and his/her experience in Italy:

```
</dm>
<dm n="902" sp-act="acknowledge">
```

```
mhm
</dm>
<dm n="903" sp-act="init">
so
</dm>
<frag n="904" sp-act="" topic="location" mode="decl">
youre happy wi with with the fact that you are here <punc
type="stop" />
</frag>
</turn>
<turn n="527" speaker="b">
<yes n="905" sp-act="acknowledge">
yes <punc type="stop" />
</yes>
```

In `frag n=904` the speech act function enclosed in the question "youre happy wi with the fact that you are here" is undoubtedly a request for information on a state of things, as was identified by all intercultural mediators and which can be explicated as follows:

```
</dm>
<dm n="902" sp-act="acknowledge">
mhm
</dm>
<dm n="903" sp-act="init">
so
</dm>
<frag n="904" sp-act="reqInfo" topic="location" mode="decl">
youre happy wi with with the fact that you are here <punc
type="stop" />
</frag>
</turn>
<turn n="527" speaker="b">
<yes n="905" sp-act="acknowledge">
yes <punc type="stop" />
</yes>
```

The speech act function attribution which was carried out manually after collecting all the information provided by the intercultural mediators involved in the analysis has enabled us to improve – albeit to some extent – the final annotated corpus, whose accurate version shall also allow researchers – once the annotated corpus has been made available online – to conduct research which does not merely rely on automated processes of speech act definition and attribution but also on a data set that is somewhat qualitatively assessed and annotated.

6. Conclusions

The present study has aimed at providing some insights into the possible applications of (semi-)automated means for speech act function retrieval and attribution. More specifically, we focused on the *DART* software tool for the annotation of speech acts in a corpus of conversation in ELF in asylum-seeking contexts. As a bedrock for our analysis we adopted a methodology that combined the fundamentals of corpus linguistics and corpus pragmatics with the most recent techniques of discourse annotation. The two case studies provided in the sections above have revealed the extent to which speech acts cannot always be automatically retrieved by means of automated software tools, but are rather context-sensitive and in most cases undergo − as is the case of other grammatical aspects of discourse in ELF, e.g. conjunctions − a process of 're-semanticization' (Centonze 2013), by means of which certain aspects of both spoken and written registers tend to overlap, negotiate a new meaning or simply become hybridized forms. The retrospective verbal report phase allowed us to compensate for this lack of accurateness on the part of the software tool that was adopted for the purposes of our study. Certainly, such an approach is experimental and much is yet to be done in order to generalize findings. Notwithstanding this, such an approach could start to be adopted in several domains and, most of all, in those multicultural contexts which see the intercultural mediator acting as an interpreter among people belonging to different socio-cultural backgrounds. Constructing a corpus and implementing it would allow a more in-depth analysis of different aspects of both spoken and written discourse in ELF and, with special reference to *DART*, a better understanding of how meaning is negotiated through the use of speech acts in spontaneous/semi-spontaneous discourse. Training intercultural mediators in this sense would become necessary and research carried out in this field would undoubtedly provide some useful insights into the dynamics of ELF in multicultural contexts.

Bionote: Laura Centonze is a PhD candidate in English Applied Linguistics at the Università del Salento and the University of Vienna. In the past few years, she was a Lecturer in English Language and Translation for the degree courses in Engineering, Archaeology and Economics. She graduated with honours in Literary Translation in 2011 and holds two post-graduate masters, respectively in European Translators (Vistula University, Poland) and in Intercultural Linguistic Mediation in Asylum-seeking Contexts (Università del Salento). Her main interests include intercultural mediation, corpus pragmatics and the exploration of new research methods for the analysis of spoken discourse in ELF in migration encounters.

Author's address: laura.centonze@unisalento.it

References

Aijmer K. and Rühlemann C. (eds.) 2015, *Corpus pragmatics. A handbook*, Cambridge University Press, Cambridge.

Austin J.L. 1962, *How To Do Things With Words*, Clarendon Press, Oxford.

Centonze L. 2013, *Conjunctions in ELF academic discourse: a corpus-based analysis*, in "Lingue e Linguaggi" 10 [2], pp. 7-18.

Centonze L. forthcoming, *Using discourse annotation software to explore English as a Lingua Franca in Migration Encounters. A corpus-pragmatic approach,* PhD Thesis, Lecce/Vienna, Università del Salento/Universität Wien.

Cogo A., Archibald A. and Jenkins J. (eds.) 2011, *Latest trends in ELF research,* Cambridge Scholars Publishing, Cambridge.

Ericsson K.A. and Simon H.A. 1984, *Protocol analysis: Verbal reports as data,* The MIT Press, Cambridge, MA.

Mauranen A. and Ranta E. 2008, *English as an Academic Lingua Franca - The ELFA Project*, in "Nordic Journal of English Studies" 7 [3], pp. 199-202.

Searle J.R. 1969, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, Cambridge.

Seidlhofer B. 2001, *Closing a Conceptual Gap: The Case for a Description of English as a Lingua Franca*, in "International Journal of Applied Linguistics" 11 [2], pp. 133-58.

Seidlhofer B. et al. 2013, *VOICE The Vienna-Oxford International Corpus of English* (version 2.0 online). Director: Barbara Seidlhofer; Researchers: Angelika Breiteneder, Theresa Klimpfinger, Stefan Majewski, Ruth Osimk, Marie-Luise Pitzl. http://voice.univie.ac.at/ (20/02/17).

Weisser M. 2014, *Pragmatic Annotation and Analysis in DART*. https://www.academia.edu/8444880/Pragmatic_Annotation_and_Analysis_in_DART (20/02/17).

Weisser M. 2015, *DART – Dialogue Annotation Research Tool*, 32bit Windows. http://martinweisser.org/ling_soft.html#DART (20/02/17).

Weisser M. 2016, *The DART speech-act taxonomy version 1.1.* http://martinweisser.org/publications/DART_taxonomy_v1.1.pdf

Widdowson H.G. 1996a, *Linguistics,* Oxford University Press, Oxford.

Appendix

Speech act categories in *DART* v. 1.1 (re-adapted from Weisser 2016)
(http://martinweisser.org/publications/DART_taxonomy_v1.1.pdf)

Speech-act Label	**(Approximate) Function**
abandon	abandoning a unit, either choosing not to complete it or due to interruption
accept	responding in an active positive way
acknowledge	signalling decoding, understanding
add	signalling extension/elaboration of information
agree	signalling explicit agreement
answer	answering a question
apologise	apologising
approve	expressing appreciation or approval
attribute	expressing attribution to s.o.
bye	saying farewell; closing a dialogue
complete	completing the interlocutor's move
conclude	indicating a (logical) conclusion
contrast	indicating a contrast, e.g. by means of a contrastive conjunction
confirm	confirming a request for confirmation
correct	correcting what the interlocutor has said
correctSelf	correcting one's own utterance
direct	eliciting the interlocutor's non-verbal response
echo	repeating the interlocutor's words for verification
elab	elaborating the answer to a question or a directive
enumerate	enumerating
exclaim	expressing emotion or surprise
explain	providing an explanation
expressAwareness	expressing awareness, possibly knowledge of s.th.
expressNonAwareness	negative counterpart to the above
expressConviction	expressing conviction, e.g. through use of *of course*
expressOpinion	expressing an opinion/evaluation
expressPossibility	expressing a possibility
expressImPossibility	negative counterpart to the above
expressRegret	expressing regret
expressStance	expressing one's attitude, e.g. through *frankly (speaking)*
expressSurprise	expressing surprise
expressWish	expressing a wish or desire
greet	greeting the interlocutor
hesitate	hesitating before the beginning of a turn/unit
hold	signalling to the interlocutor to hold the line, usually to look up information or to think
identifySelf	identifying the speaker's name/institution
init	initiating a new phase of the dialog
insult	insulting the interlocutor

negate	responding negatively
offer	offering a service to benefit the interlocutor
pardon	signalling misunderstanding/the need for the interlocutor to repeat
phatic	semantically empty discourse-marking expression, such as initial *you know*
predict	predicting some future event
predictPossibility	predicting a possibility
promise	making a promise
refer	indicating a deictic reference (neutral option)
referCondition	referring to a condition
referOpt	referring to an option
referPerson	referring to a person (excluding vocatives)
referReason	referring to a reason
referTime	referring to a specific (point in) time
referThing	referring to a concrete or abstract object
refuse	responding negatively to an offer, etc.
reject	rejecting a proposal
reqConfirm	requesting a confirmation
reqDirect	requesting a directive
reqInfo	requesting verbal information
reqModal	requesting permission, advice, etc.
reqOpt	requesting an option
selfTalk	speaking to oneself (the speaker)
spell	spelling out something
state	conveying information/awareness
stateIntent	indicating the speaker's intention
stateConstraint	stating a potential constraint
stateOpt	stating a potential option
stateReason	stating a reason
summarise	signalling a summary
suggest	proposing action by the interlocutor (or the interlocutor and the speaker)
suggestOpt	suggesting a potential option
swear	expressing an expletive
thirdParty	speaking to s.o. who is not the speaker or the interlocutor
thank	thanking
unclassifiable	a speech-act not classifiable according to the present scheme
uninterpretable	uninterpretable, due to missing or incoherent information