

From corpora to specialized semantic frames

Ana Ostroški Anić

Martina Pavić

Institute for Croatian Language and Linguistics

- project the *Dynamicity of Specialized Knowledge Categories* (DIKA) ihjj.hr/dika
- head of the project: Ana Ostroški Anić
- funded by the Croatian Science Foundation
- 10 March 2018 – 9 March 2023
- 6 researchers and 1 doctoral student

Project goals

- theoretical research of terminological units and specialized knowledge categories at conceptual and linguistic levels
- translator-oriented descriptive terminological database
- phraseological terminological units and the role of figurativeness in the construction and expression of specialized meaning
- syntactic and semantic analyses of the domains of aviation, marketing, law and carstology

Anticipated results

- parallel corpus in the domain of air traffic ✓
- contrastive analyses of the translation equivalents in Croatian, English, German and French
- terminological database of semantic frames *AirFrame* – organization of the fundamental semantic frames from the field of aviation (Flight, Airplane, Airspace, Airport, Air_traffic and Air_traffic_control)

Aim

- first steps in developing a method for determining semantic frames and their conceptual relations
- parallel corpus used for the extraction of terminological units referring to specific semantic frames of aviation

Theoretical background

- over the past two decades focus of terminological research has shifted to processes and events in a specialized domain
- predicative terms (verbs, adjectives, adverbs and certain nouns) often denote key concepts of a domain terminology: e.g. *communicate*, *fly*, *take off*, and *transport* in aviation
- criteria for defining types of verbs carrying specialized meaning proposed by L'Homme (1998, 2014) most referred to in terminological literature
 - a) verbs specific to a certain field (*land*, *take off*), b) verbs that acquire specialized meaning in the field (*operate*), and c) verbs of general language.

Methods

- 1) parallel English-Croatian corpus was created in *Sketch Engine*
- 2) term extraction and validation of term candidates
- 3) extracting a list of verbs and analyzing their collocations bearing specialized meaning

Corpus compilation

- the corpus is compiled from the Directory of legal acts of the European Union, chapter *Transport policy*, subchapter *Air transport* in English and Croatian
- out of 220 documents from the *Air transport* subchapter, 178 legal acts are taken having both language versions

	English	Croatian
Tokens	1,151,297	1,059,406
Words	951,156	855,56
Sentences	72,045	75,638
Documents	178	178

PARALLEL CONCORDANCE

air transport, English



Get more space +

simple fly 346 (300.53 per million)

align



zračni prijevoz, Croatian

<s> In the case of commercial air transport, the 90-day period prescribed in subparagraphs (b)(1) and (2) above may be extended up to a maximum of 120 days, as long as the pilot undertakes line flying under the supervision of a type rating instructor or examiner. </s>	<s> U slučaju komercijalnog zračnog prijevoza, razdoblje od 90 dana propisano u (b)(1) i (2) gore smije se produžiti do maksimalno 120 dana, ako pilot obavi linijsko letenje pod nadzorom instruktora za tip ili ispitivača. </s>
<s> perform the additional flight time or take-offs and landings, flying dual or solo under the supervision of an instructor, in order to fulfil the requirements in (a). </s>	<s> odraditi dodatni nalet ili polijetanja i slijetanja, s instruktorom ili samostalno pod nadzorom instruktora, u svrhu ispunjavanja zahtjeva iz (a). </s>
<s> perform the additional flight time or take-offs and landings, flying dual or solo under the supervision of an instructor, in order to fulfil the requirements in (a). </s>	<s> obaviti dodatno letenje ili polijetanja i slijetanja, s instruktorom ili samostalno pod nadzorom instruktora, u svrhu ispunjavanja zahtjeva iz (a). </s>
<s> (d) When the pilot does not comply with the requirement in (c), he/she shall perform the additional number of launches flying dual or solo under the supervision of an instructor in order to renew the privileges. </s>	<s> (d) Kad pilot ne udovoljava zahtjevima iz (c), mora obaviti dodatni broj startova s instruktorom ili samostalno pod nadzorom instruktora u svrhu obnavljanja privilegija. </s>
<s> When the holder of the LAPL(S) also has the privileges to fly aeroplanes, the requirements in (1) may be completed on aeroplanes. </s>	<s> Kad imatelj LAPL(S) ima i privilegije za letenje aviona, zahtjev iz (1) može se obaviti i na avionima. </s>
<s> perform the additional flight time or take-offs and landings, flying dual or solo under the supervision of an instructor, in order to fulfil the requirements in (a) or (b). </s>	<s> obaviti dodatni nalet ili polijetanja i slijetanja, s instruktorom ili samostalno pod nadzorom instruktora, u svrhu ispunjavanja zahtjeva iz (a) ili (b). </s>
<s> (d) When the pilot does not comply with the requirement in (c), he/she shall perform the additional number of tethered flights flying dual or solo under the supervision of an instructor in order to renew the privileges. </s>	<s> (d) Kad pilot ne udovoljava zahtjevima iz (c), mora obaviti dodatan broj letova balonom sa sponom s instruktorom ili samostalno pod nadzorom instruktora u svrhu obnavljanja privilegija. </s>
<s> in addition, if the pilot is qualified to fly more than one class of balloons, in order to exercise their privileges in the other class, they shall have completed at least 3 hours of flight time in that class within the last 24 months, including 3 take-offs and landings. </s>	<s> dodatno, ako je pilot kvalificiran za letenje s više od jedne klase balona, a da bi mogao koristiti privilegije na tim drugim klasama balona, mora na svakoj klasi u prethodnih 24 mjeseca obaviti 3 sata naleta uključujući 3 polijetanja i slijetanja. </s>
<s> perform the additional flight time or take-offs and landings, flying dual or solo under the supervision of an instructor, in order to fulfil the requirements in (a). </s>	<s> obaviti dodatni nalet ili polijetanja i slijetanja, s instruktorom ili samostalno pod nadzorom instruktora, u svrhu udovoljavanju zahtjeva iz (a). </s>

Term extraction and verification

- automatic term extraction was conducted for each language
 - EUR-Lex English 2/2016 corpus and EUR-Lex Croatian 2/2016 served as reference corpora for single-word lists; English Web 2013 and Croatian Web hrWaC 2.2. reference corpora for multi-word keywords
- verification process of candidates' lists done in two phases
 1. removing all abbreviations and general language words and collocations
 2. analyzing concordances of remaining terms

Extraction and analysis of verbs

- lists of English and Croatian verbs from respective corpora extracted automatically using the Keywords option in *Sketch Engine*
- the first 100 verbs from each list were then manually validated through their word sketches
- collocations determined as multi-word terms or collocations of relative significance for aviation terminology noted down in an Excel sheet next to the analyzed verb

Results

- out of 1000 single-word candidates from the Croatian corpus, 199 terms labeled as single-word terms; 455 multi-word terms
- 664 multi-word English terms remained after the analysis, only 148 single-word
- poor result of only 14.8 % valid terms from the initial list of multi-word English terms

Discussion

- TERM VERIFICATION posed several issues
- if both the noun and the adjective referring to a certain concept appeared on the list, the noun was kept as a more prototypical term realization $\Rightarrow$ more common in Croatian
 - in English a noun and an adjective often share the morphological form (*kabina*, *kabinski* 'cabin', *identifikacija*, *identifikacijski* 'identification')
- when two nouns referring to the same concept appeared, both were kept if one of them was referring to the **action**, and the other to the **result**: *identifikacija* and *identificiranje* (Eng. identification)

Discussion (ctd.)

- false positive multi-word terms

1. domain knowledge required

bound air > bound air cargo

simulated engine > simulated engine failure

visual flight > visual flight rules

2. collocations of terms

relevant aircraft, individual aircraft, specific aircraft, large aircraft

entire flying time

Verb analysis

analysis of extracted verbs – two steps

- 1) specialized verbs or verbs specific to aviation were marked
- 2) marking verbs that potentially acquire a specialized meaning in aviation context
- first 100 most frequent verbs were analyzed (word sketches)
- the largest number of verbs included support verbs and phraseological verbs

e.g. *provide, ensure, require* or *refer to* gain specialized meaning in syntagmatic units > *establish aerodrome, ensure airworthiness*

operate as verb 1,564x

modifiers of "operate"	objects of "operate"	subjects of "operate"	"operate" and/or ...	prepositional phrases	pronominal subjects of "operate"	infinitive objects of "operate"	-ing objects of "operate"
actually ... actually operating the flight	ban ... subject to an operating ban	engine ... with all engines operating	intend ... carriers operating or intending to operate at	"operate" in "operate" under "operate" by "operate" at "operate" on "operate" as "operate" within "operate" with "operate" during "operate" between "operate" over "operate" into ▼	they ... they operate it ... agents) ; it is operated so as to	serve ... it is operated so as to serve traffic between the	use ... maintained and operated using the appropriate and
independently ... transponder signals which operates independently of ground-based equipment	licence ... a valid operating licence	carrier ... carriers operating	provide ... organisations operating or providing services at the				
currently ... types or variants currently operated . The	restriction ... noise-related operating restrictions	aircraft ... aircraft operating in	maintain ... built , maintained and operated using the appropriate				
only ... shall only operate an aircraft	site ... aerodrome or operating site	flight ... flights operating	manufacture ... designed , manufactured , operated				
as ... it is operated so as to serve traffic	carrier ... the operating air carrier	hems ... at a HEMS operating site	manage ... Agency shall manage and operate the tools and				
not ... shall not operate a	flight ... operating flights	helicopter ... a helicopter operating under					
	procedure ... working methods and operating procedures	airline ... airlines operating					
	aircraft ... aircraft operated	aerodrome ... aerodrome operating minima					
	authorisation ... respect of the operating authorisation of that air	organisation ... organisations operating or providing services					
	helicopter ... Helicopters operated	aeroplane ... Aeroplanes operated					
	aeroplane ... Aeroplanes operated	variant ... aircraft types and variants operated					
	service ... operate the data link services	vendor ... vendors operating in the					

Verb analysis (ctd.)

- a verb is likely to be a term if “the linguistic realizations of actants (i.e. arguments) of a verb are defined as terms” (L’Homme 2014)
- in the first 100 most frequent verbs in the English part of the corpus, only two verbs can be defined as proper specialized verbs in aviation – *operate* and *fly*
- in the Croatian corpus only *letjeti* ‘fly’ appears as a proper specialized verb

Conclusions

- terminological and linguistic data extracted from specialized parallel corpora can be used in setting the foundation for defining specialized semantic frames
- term validation has shown that the majority of proposed terms could not be classified as aviation related terms
- collocations of verbs specific to the field of aviation are terminologically relevant
- using verbs as a starting point in defining specialized semantic frames of aviation might be less useful than analyzing word sketches
- the next step – grouping relevant concepts and naming the frames they invoke
- improved automatic definition extraction would get better results for Croatian.

Thank you!

aostrosk@ihjj.hr
mpavic@ihjj.hr

This work has been fully supported by the Croatian Science Foundation under the project HRZZ-UIP-2017-05-7169.