

From concept definitions to semantic role labeling in specialized knowledge resources

Ivana Brač and Ana Ostroški Anić

Institute of Croatian Language and Linguistics

INTRODUCTION

Struna is a Croatian terminological database with the terminology of 29 subject fields. It currently contains 46 categories, 4 of which are mandatory (preferred term, grammatical information, definition, and an equivalent in English).

AirFrame is a single domain terminological resource. It is a lexical database in which aviation concepts and their definitions, as well as terms in several languages, and other relevant categories of terminological entries are presented in specialized semantic frames.

The database will include these categories:

- semantic frames
- frame elements and their definitions
- terms and terminological units (collocations and phraseological units – corresponding to lexical units in FrameNet)
- sentences labeled with semantic roles
- frame-to-frame conceptual relations.

METHODOLOGY

Definitions of key aviation concepts as defined in Struna were analyzed and compared to sentences containing terms for those concepts, in order to establish the difference in conceptual and semantic information that can be gathered from terminological definitions as opposed to annotated examples from corpora. The list of key aviation concepts used as target words for extracting sentences included: *aerodrome*, *aircraft*, *airline*, *airspace*, *air traffic control*, *air transport*, *flight*, *landing* and *taking off*.

Definitions consist of the superordinate or a more generic concept (to the one that is being defined), and the delimiting characteristics, i.e. the characteristics that set out the defined concept from concepts similar or related to it.

let	im. m. jd.
definicija	odlazak s određenoga aerodroma prema određenoj određenoj aerodromu
istovrijednice	engleski: flight
razredba	polje: tehnologija prometa i transport grana: zračni promet projekt: Ustavljanje i normiranje temeljnoga hrvatskoga zrakoplovnog nazivja

Figure 1. Definition of the term *flight* (Struna)

TERM	DEFINITION
aerodrome	a defined area on land or water, including any objects , installations and equipment , intended to be used either wholly or in part for the movement, take-off, landing and parking of aircraft
airline	an air operator that uses aircraft to transport people and/or goods for commercial purposes
air transport	the transport of people or things from one place to another by means of an aircraft
flight	the flying of aircraft from any aerodrome to a destination aerodrome
take-off	an aircraft operation during which an aircraft accelerates from the stop phase , leaves the ground and reaches the required flight level

Figure 2. Definitions in which frame elements are marked.

A parallel English-Croatian aviation corpus compiled within the DIKA project was queried for English and Croatian terms of the analyzed aviation concepts. The corpus is compiled from the Directory of legal acts of the European Union, chapter Transport policy, subchapter Air transport. Out of 220 documents, 178 legal acts are taken having both language versions. The texts are downloaded from the EUR-Lex database, and entered into the Sketch Engine's corpus compilation module (Kilgariff et al., 2014).

<> Traffic on a given flight with the same flight number subdivided by airport pairs in accordance with point of embarkation and point of disembarkation on that flight. </>	<> Promet na određenom letu s istim brojem leta podijeljen između parova zračnih luka u skladu s točkom iskrcaja na tom letu. </>
<> An air transport flight or series of flights for the public transport of passengers and/or freight and mail, for remuneration or for hire. </>	<> Let ili niz letova u zračnom prijevozu za javni prijevoz putnika i tereta i pošte za naplatu ili drugo. </>
<> On some flights with intermediate stops, the flight number changes at an airport to designate the change between an inbound and outbound flight. </>	<> Na pojedinim letovima s međuslijetanjima, broj leta se mijenja u zračnoj luci kako bi se označila leta u dolasku i leta u odlasku. </>
<> (13) Passengers whose flights are cancelled should be able either to obtain reimbursement of their tickets or to obtain re-routing under satisfactory conditions, and should be adequately cared for while awaiting a later flight . </>	<> Putnici čiji je let otkazan trebali bi moći ili ostvariti povrat novca vrijednosti karte ili ostvariti preusmjeravanje u drugim uvjetima i trebali bi biti zbrinuti na odgovarajući način dok čekaju kasniji let. </>
<> (h) "final destination" means the destination on the ticket presented at the check-in counter or, in the case of directly connecting flights, the destination of the last flight ; alternative connecting flights available shall not be taken into account if the original planned arrival time is respected. </>	<> "konačno odredište" znači destinacija na karti predstavljena na šalteru za registraciju putnika ili, u povezanosti letova , odredište zadnjeg leta; zamjenski povezani letovi koji su dostupni neće uzimati u planiranju vrijeme dolaska. </>
<> When passengers are offered re-routing to their final destination on an alternative flight pursuant to Article 8, the arrival time of which does not exceed the scheduled arrival time of the flight originally booked. </>	<> Kada je putnicima ponudeno preusmjeravanje do njihovog mjesta odredišta zamjenskim letom , i kad na taj način ostvareno vrijeme dolaska ne premašuje prvotno vrijeme dolaska leta. </>
<> (b) use of available airspace capacity during all phases of flight , including slot assignment, and </>	<> korištenje raspoloživog kapaciteta zračnog prostora tijekom svih faza leta , uključujući dodjelu slotova </>
<> Flight data processing systems shall be interoperable in terms of the timely sharing of correct and consistent information, and a common operational understanding of that information, in order to ensure a coherent and consistent planning process and resource-efficient tactical coordination throughout the EATMN during all phases of flight. </>	<> Sustavi za obradu podataka o letu moraju biti interoperabilni u smislu pravodobne razmjene točnih informacija te njihova zajedničkog operativnog tumačenja od strane svih sudionika kako bi se osiguralo dosljedan postupak planiranja i taktička koordinacija s učinkovitom uporabom resursa u cijelom EATMN faza leta. </>
<> The provisions of this Chapter shall apply whether the flight is scheduled or non-scheduled and whether the flight is part of a package or not. </>	<> Odredbe ovog poglavlja primjenjuju se bez obzira na to da li je let redovit ili ne, i da li je let dio paketa ili ne. </>
<> "flight intention" means the flight path and associated flight data describing the planned trajectory of a flight to its destination, as updated at any moment. </>	<> "namijena leta " znači letna putanja i povezani podaci o letu koji opisuju planiranu putanju leta do se ažurira u svakom trenutku. </>

Figure 3. Parallel concordances of the English term *flight* and its Croatian equivalent *let*.

The corpus was queried for the target terms *aerodrome*, *aircraft*, *airline*, *airspace*, *air traffic*, *air traffic control*, *air transport*, *flight*, *landing* and *take-off*, and sentences were manually extracted and annotated. Extracted sentences were annotated for semantic roles using the LIRICS semantic tagset (Petukhova & Bunt, 2008), while the annotation was carried out using the WebAnno tool (Eckart de Castilho et al., 2016).

RESULTS

Since traditional terminology work is largely focused on defining entities as prototypical categories for knowledge representation, it was expected that entities would comprise the largest share of the types of categories labeled as frame elements in definitions. Apart from entities such as *aircraft*, *airplane*, *person* or *thing*, bounded regions like *place*, *area* and *aerodrome areas* also fall within a large group of entities, but are labeled as locations according to the terminology of FrameNet. Activities like *flying*, *take-off*, *movement* and *transport* are typically present in the field of aviation, as well as a number of procedures that we ontologically define as a type of an activity. Ontological categories labeled in the analyzed definitions refer to frame elements that invoke particular semantic frames that structure the field of aviation.

Figure 4. Annotated frame elements as represented in Croatian definitions extracted from Struna

Decisions had to be made whether to:

1. use frame elements - an exhaustive list of roles, but a difficulty to correlate with resources that apply a more general and limited SRL tagset
2. customize own list of semantic roles
3. accept one of the existing - e.g. LIRICS, pro - correlate with other resources using it, con - lost of some useful semantic information (duration)

The flight previously notified by a basic flight data process [THEME] will now not enter the airspace of the notified unit [FINAL_LOC].
An applicant [AGENT] shall fly the aircraft [THEME] from a position where the PIC functions can be performed [INITIAL_LOC] and to carry out the test [THEME] as if there is no other crew member [SETTING].
An aircraft taxiing on the manoeuvring area of an aerodrome [AGENT] shall give way to aircraft taking off or about to take off [BENEFICIARY].
Similarly, the airline operating the aircraft [PIVOT] needs underlying economic authority [THEME] from the DOT [SOURCE].

Figure 5. Semantically annotated example sentences.

Categories of specialized knowledge can be defined in terms of semantic relations that bind them together, but it must nevertheless be done in a precise, clear and concise manner so to reflect both terminological needs and the need for good terminology.

REFERENCES

- Brač, I., M. Bratanić, & A. Ostroški Anić. (2015). Hrvatsko nazivlje i nazivoslovlje do Šuleka do Strune - hrvatski jezik i terminološko planiranje. In M. Bratanić, I. Brač, & B. Pritchard (Eds.), *Od Šuleka do Schengena: terminološki, terminografski i prijevodni aspekti jezika struke* (pp. 3-26). Zagreb/Rijeka: Institut za hrvatski jezik i jezikoslovlje & Pomorski fakultet u Rijeci.
- Bratanić, M., & A. Ostroški Anić. (2013). The Croatian National Termbank STRUNA: A New Platform for Terminological Work. *Collegium Antropologicum*, 37(3), 677-683.
- Eckart de Castilho, R. et al. (2016). A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, Osaka, Japan (pp. 76-84). Retrieved from <https://pdfs.semanticscholar.org/ae7/1bfa28ec1bad78d4bd4aadcab168aa6b3b13.pdf>
- Kilgariff, A. et al. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7-36.
- Petukhova, V., & H. Bunt. (2008). LIRICS semantic role annotation: design and evaluation of a set of data categories. *Proceedings of the Sixth International Conference on Language Resources and Evaluation* (pp. 39-45). Retrieved from <https://pdfs.semanticscholar.org/732c/65885e1e664c44db6ad723425547633dad7a.pdf>
- Ruppenhofer, J. et al. (2010). *FrameNet II: Extended Theory and Practice*. Retrieved from <https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>