Superframes for Consistent and Comprehensive Semantic Role Annotation

Kilian Evang

Heinrich Heine University Düsseldorf
Germany
evang@hhu.de

1 Introduction

We present CRANS, a new annotation scheme for comprehensively labeling syntactic dependencies from predicates to dependents in running text with their semantic functions, or *roles*. CRANS is designed to be applied atop UD (de Marneffe et al., 2021), SUD (Gerdes et al., 2018), or EUD (Schuster and Manning, 2016), explicitly annotating the syntax-semantics interface rather than just semantics. An example of a UD tree enriched with CRANS annotations is shown in Figure 1. Although we use English examples in this abstract, CRANS is designed to be applied cross-linguistically. The vocabulary of CRANS consists of:

Core Roles   There are only two core roles in CRANS: F (figure) and G (ground). A predicate that has any arguments must always have at least one core argument. This rule is meant to ensure consistent annotation and high inter-annotator agreement. What F and G mean exactly depends on the *superframe* of the predicate.

(1) a. Kim$_G$ owns$_{PSS}$ a house$_F$
    b. Kim$_F$ married$_{SOC}$ Sandy$_G$

Superframes   Each predicate instance is labeled with one of a handful of coarse frames, called *superframes*. CRANS aims to distinguish only as many superframes as necessary to give a clear mapping between arguments and core roles. Table 1 lists each superframe along with its definitions for F and G.

Derived Roles   In addition to F and G, there is IG (initial ground). This is used for arguments that participate in the G role at the beginning of the scene and that either continue to do so or move away from this state. There is also the MG role for grounds that are "passed through".

(2) a. Kim$_{IG}$ kept$_{PSS}$ the money$_F$
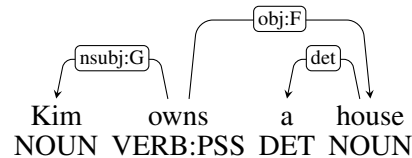   b. Kim$_{IG}$ lost$_{PSS}$ the money$_F$

Figure 1: UD tree enriched with CRANS semantic role annotation

c. Kim$_F$ divorced$_{SOC}$ Sandy$_{IG}$
d. Kim$_{IG}$ gave$_{PSS}$ Sandy$_G$ the house$_F$
e. The noise$_{IG}$ faded$_{SCN}$
f. Kim$_F$ went$_{CNF}$ from the living room$_{IG}$ through the door$_{MG}$ into the kitchen$_G$

Modifier Roles   CRAN's inventory of modifier roles is not much different from other schemes. Modifiers are dependents that are not selected for by the predicate, but can freely combine with all kinds of predicates. A preliminary list of modifier roles is given in Table 2.

(3) Kim$_F$ went$_{CNF}$ to Brazil$_G$ last month$_{TMP}$ because they needed$_{XPL}$ a vacation

Mix-in Roles   Arguments that are selected for by the predicate but are not core arguments are labeled with mix-in roles. These are just modifier roles prefixed with an X. Very common mix-in roles are XCAU (causer), XSND (sender), and XRCP (recipient).

(4) a. Kim$_{XCAU}$ broke$_{STP}$ the vase$_F$
    b. Kim$_{XSND}$ talked$_{MSG}$ shit$_G$ about Sandy$_F$ to Aubrey$_{XRCP}$
    c. Kim$_{XRCP}$ saw$_{MSG}$ Sandy$_F$ swim$_G$
    d. Kim$_{XRCP}$ searched$_{MSG}$ the woods$_{XLOC}$ for Sandy$_F$
    e. Kim$_{IG}$ paid$_{PSS}$ a million dollars$_F$ for the house$_{XAST}$

Aktionsart   Apart from the IG/MG/G distinction, superframes abstract away from aktionsart. For example, states have the same superframes and roles as the events bringing them about, and attempts

Superframe	Description	F	G
WST	world state		
ACT	activity	actor	
EXP	experience	undergoer	
STP	state/property	entity in state/with property	
ASS	assignment	point	value
CMP	comparison	compared	reference
SUC	succession	successor	succeeded
QTY	quantity	of what	how much
CNF	configuration	smaller/peripheral entity	larger/central entity
ORL	organizational role	appointee	organization/job/responsibility
PSS	possession/control	possession	possessor
PWH	part-whole	part	whole
SOC	social	person	friend, relative, etc.
SCN	scene	participant	scene
MSG	message	topic	content

Table 1: Meaning of the F (figure) and G (ground) core roles depending on the superframe. When the semantic criteria do not distinguish F and G, F is the syntactically less oblique dependent.

and failures have the same superframes and roles as successes.

(5) a. Kim$_G$ owns$_{PSS}$ a house$_F$
    b. Kim$_G$ bought$_{PSS}$ a house$_F$

(6) a. Kim$_{XRCP}$ remembered$_{MSG}$ Sandy$_F$
    b. Kim$_{XRCP}$ forgot$_{MSG}$ Sandy$_F$
    c. Kim$_F$ tried$_{SCN}$ to sleep$_G$
    d. Kim$_F$ managed$_{SCN}$ to sleep$_G$
    e. Kim$_F$ failed$_{SCN}$ to sleep$_G$

Etc. We have created detailed guidelines for annotating less prototypical examples, including: nominal and adjectival predicates; non-roles for expletives and extended nuclei in multiword expressions; a uniform treatment of auxiliaries, light verbs, raising and control verbs; additional annotation of argument-argument relations such as control; and dual framing in case of uncertainty, secondary predicates, and idiomatic language.

2 Comparison

CRANS vs. other Frame-based SRL Schemes
VerbNet (Kipper et al., 2008), FrameNet (Fillmore and Baker, 2009), PropBank (Palmer et al., 2005), and VerbAtlas (Di Fabio et al., 2019) all presuppose a (relatively) large frame lexicon. In our experience, lexicons have the problem of being perennially incomplete, and taking up much time to browse during annotation. VerbAtlas, like CRANS, makes an effort to limit the number of frames by grouping predicates together into frames with coherent rolesets. However, at 433 frames, it still requires much looking up. It also has a number of inconsistencies in role names and PropBank mappings, stemming from the semi-automatic creation process. CRANS aims to trade the rich ontology that especially FrameNet provides for ease and speed of annotation.

CRANS vs. Frameless SRL Schemes Semantic role vocabularies can be used without (explicit) reference to frames. For example, the Parallel Meaning Bank (Abzianidze et al., 2017) uses a variant of the VerbNet role inventory without frames. Frame-specific numbered PropBank roles are also all annotated with "function tags", i.e., frame-independent role labels. In our experience, it is hard to apply such a scheme consistently, as it is often hard to decide whether something is an Agent or a Theme, or a Patient or a Theme or a Topic, without having a frame that specifies that. We feel that by forcing annotators to choose a superframe and defining the core roles and their relation to each other in each superframe, CRANS provides a better handle on choosing roles. Although also frameless, the adposition-focused SRL scheme SNACS (Schneider et al., 2018; Shalev et al., 2019) has strongly

Scene	
AST	asset
BEN	beneficiary
CAU	causer
EXT	extent
INS	instrument
LOC	locus
MNR	manner
TMP	temporal
Discourse	
CNC	concession
CNT	continuation
CTX	context
ELA	elaboration
RCP	recipient
SND	sender
XPL	explanation
Constructional	
ANC	ancillary
ATT	attribute
DPC	depictive
RSF	result: affected entity
RSG	result: end state

Table 2: Modifier roles, roughly divided into modifiers that specify further properties of a scene (Scene), those relating it to other scenes (Discourse), and Constructional ones giving rise to argument-argument semantic dependencies.

inspired CRANS's inventory of superframes and modifier roles. The most important difference is that we removed argument roles like Agent, Theme, or Source, which we see as ill-defined, and introduced roles like F and IG, which are well-defined within their superframes.

3 Limitations

CRANS is in an early stage of development and has yet to be evaluated in annotation experiments. With this poster, we would like to gather feedback and meet potential collaborators for such an endeavor.

Acknowledgements

References

Lasha Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik van Noord, Pierre Ludmann, Duc-Duy Nguyen, and Johan Bos. 2017. The Parallel Meaning Bank: Towards a multilingual corpus of translations annotated with compositional meaning representations. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 242–247, Valencia, Spain. Association for Computational Linguistics.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Andrea Di Fabio, Simone Conia, and Roberto Navigli. 2019. VerbAtlas: a novel large-scale verbal semantic resource and its application to semantic role labeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 627–637, Hong Kong, China. Association for Computational Linguistics.

Charles J. Fillmore and Collin Baker. 2009. A frames approach to semantic analysis. In Bernd Heine and Heiko Narrog, editors, *The Oxford Handbook of Linguistic Analysis*, pages 791–816. Oxford University Press, Oxford, UK.

Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. 2018. SUD or surface-syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 66–74, Brussels, Belgium. Association for Computational Linguistics.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation*, 42:21–40.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Nathan Schneider, Jena D. Hwang, Vivek Srikumar, Jakob Prange, Austin Blodgett, Sarah R. Moeller, Aviram Stern, Adi Bitan, and Omri Abend. 2018. Comprehensive supersense disambiguation of English prepositions and possessives. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 185–196, Melbourne, Australia. Association for Computational Linguistics.

Sebastian Schuster and Christopher D. Manning. 2016. Enhanced English Universal Dependencies: An improved representation for natural language understanding tasks. In *Proceedings of the Tenth International Conference on Language Resources and*

Evaluation (LREC'16), pages 2371–2378, Portorož, Slovenia. European Language Resources Association (ELRA).

Adi Shalev, Jena D. Hwang, Nathan Schneider, Vivek Srikumar, Omri Abend, and Ari Rappoport. 2019. Preparing SNACS for subjects and objects. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 141–147, Florence, Italy. Association for Computational Linguistics.