

Framenet annotations Dutch

March 2, 2018

Contents

1	Introduction and method	2
2	Corpus composition	3
3	Annotations statistics and agreement	3
4	Frame confusions, agreement, disagreement	4
5	FN lexicon based on annotations	5
6	Other FN annotation projects	6
7	Difficult cases	8
8	Appendix I: agreement per file	10

1 Introduction and method

- Creating a Dutch corpus with FN annotations
- Full text strategy: all content words - so far only verbs - are annotated (no preselection of lexical units; no preselection of example sentences). The task was to identify the frame for a given predicate in a sentence, and assign the corresponding roles. They were provided with detailed guidelines that explained the task using multiple examples. During annotation, they had access to parsed versions of the text, and to the English FrameNet
- All frames strategy (no pre-selection of frames) : the annotator chooses the correct frame from one of the 1300+ frames from the English FN (version 1.7)
- Annotations are added on top of PropBank annotations. So, events and arguments are already identified.
- no new frames are added; if a frame could not be found the event is annotated with 'none'
- all files are double-annotated

2 Corpus composition

Table 1: Corpus statistics

theme/genre	nr_of_files	nr_of_annotated_verb_tokens
background-news	3	110
financial	17	1756
medical	1	88
news	5	499
newsletter	3	111
periodicals	37	821
policy	12	352
teletext	3	169
websites	1	49
wiki	34	1295
<i>totals</i>	116	5250

3 Annotations statistics and agreement

Table 2: FN annotation statistics and agreement

nr of verbs (tokens)=nr of frames	5250	$\kappa = 0.467$
nr of unique verbs (types)	1336	
nr of unique frames	671	
strict agreement on frames	47 %	
lenient (only inheritance relations) agreement on frames	51 %	
lenient agreement (all relations) on frames	54 %	
agreement frame elements (with matching frames)	79 %	

- strict agreement: only identical frames are considered matching (percentage agreement)
- lenient agreement (only inheritance relations): frames that are related through framenet relations 'Inherits from' or 'Is inherited by'
- lenient agreement (all relations): frames that are related through framenet relations, all relations are included
-
- agreement on frame elements: measured only when frames are matching

4 Frame confusions, agreement, disagreement

Table 3: top 20 frame confusions

19	Activity_start	Process_start
14	Creating	Intentionally_create
14	Cause_change_of_position_on_a_scale	Change_position_on_a_scale
12	Using	Using_resource
12	Opinion	Regard
10	Cooking_creation	Manufacturing
8	Getting	Receiving
8	Expressing_publicly	Statement
8	Existence	Presence
8	Awareness	Grasp
7	Operate_vehicle	Self_motion
7	Finish_competition	Finish_game
7	Causation	Evidence
7	Being_named	Name_conferral
6	Perception_active	Perception_experience
6	Intentionally_create	Text_creation
6	Giving	Grant_permission
6	Cure	Medical_intervention
6	Cause_to_perceive	Expressing_publicly
6	Beat_opponent	Finish_competition
6	Awareness	Certainty
6	Accomplishment	Getting
5	Reference_text	WrongRelation
5	Preventing	Thwarting
5	Perception_active	Reference_text
5	Have_associated	Possession
5	Finish_competition	Success_or_failure
5	Competition	Finish_competition
5	Communication	Statement
5	Communication	Expressing_publicly

Many of these confusions are a matter of specificity: using vs. using-resource; finish-competition vs. finish-game; intentionally-create vs. text-creation; beat-opponent vs. finish-competition. Others are really hard to distinguish from each other when looking at definitions and examples in FN: opinion vs. regard; getting vs. receiving; cure vs. medical-intervention. And there are also frames which are actually different and should not be confused, but which are apparently difficult to distinguish by the annotators: Activity-start vs. process-start ; create vs. intentionally-create (agentive vs. non-agentive).

Table 4: agreement on frequent frames

frame	agreements	disagreements	percentage agreement
Statement	108	60	0.64
Possession	38	64	0.37
Existence	26	62	0.3
Causation	38	45	0.46
Event	30	46	0.39
Coming_to_be	30	46	0.39
Intentionally_act	24	49	0.33
Change_position_on_a_scale	27	39	0.41
Awareness	14	50	0.22
Activity_start	23	35	0.4
Being_located	22	35	0.39
Intentionally_create	4	52	0.07
Opinion	18	38	0.32
Receiving	22	31	0.42
Inclusion	12	39	0.24
Attempt_suasion	33	18	0.65
Evidence	14	34	0.29
Cause_to_perceive	11	36	0.23
Becoming_aware	12	33	0.27
Using	17	28	0.38
None	2	40	0.05
Giving	7	35	0.17
Perception_active	17	24	0.41
Finish_competition	5	35	0.12
Self_motion	17	23	0.42
Participation	12	27	0.31
Request	19	19	0.5
Communication	6	30	0.17
Desiring	25	11	0.69
Circumscribed_existence	2	32	0.06
Process_start	8	26	0.24
Arriving	8	26	0.24
Cause_change	13	21	0.38
Removing	16	18	0.47
Accomplishment	6	27	0.18

5 FN lexicon based on annotations

We derived a FN lexicon based on the annotations collecting all lexical units or frame. (filename DFNlex.txt)

Table 5: FN annotation statistics and agreement

nr of frames	671	
nr of form units (lemmas)	1336	
nr of lexical units (frame-lemma combinations)	4755	
average nr of lexical units per frame	7	

6 Other FN annotation projects

Table 6: Other studies

[Burchardt et al., 2009] nr of unique frames nr of annotated words nr of unique words parts-of-speech agreement problems approach	German 500 20000 ? ? 0.85 (2 annos)/0.97(2 abj) metaphors, support verbs pre-selecting frames per word; automatic selection of sentences per lemma; annotating lemma-by-lemma
[Subirats, 2009] nr of unique frames nr of annotated words nr of unique words parts-of-speech agreement approach	Spanish 100 ? 600 verb, noun - pre-selecting frequent frames, pre-selecting words per frame, annotators choose sentences to annotate by searching a corpus using syntactic patterns
[Hayoun and Elhadad, 2016] nr of unique frames nr of annotated words nr of unique words parts-of-speech agreement approach	Hebrew 200 ? ? ? ? starting from 200 most freq. frames/ translations of English LUs per frame/ searching in corpus for example sentences
[Ahlberg et al., 2014] [Borin and xx, 2010] nr of unique frames nr of annotated words nr of unique words parts-of-speech agreement approach interesting extra	Swedish 1200(really annotated?) 34,000 ? ? ? ??automatic?? linked to existing LU repository
[Djemaa et al., 2016] nr of unique frames nr of annotated words nr of unique words parts-of-speech agreement approach	French 98 13,000 672 (LU instead of FU?) ? 85.9 corpus-driven 4 domains/ pre-select frames per domain/ pre-select words per frame (??)/ annotate all occurrences of a word

7 Difficult cases

Possession frame and 'hebben'

- hij heeft een ronde vorm
- hij heeft de leiding (leiden)
- hij heeft geduld (he is patient) (Experiencer_focused_emotion)
- gemeenten met meer dan 100000 aantal inwoners hebben een lik achter de naam
- de raad vaardigt decreten uit die kracht van wet hebben
- Ook halloween, concurrent van Allerzielen, heeft dezelfde oorsprong (Origin)
- Atkins beschikte over een unieke speelstijl
- hij heeft geen baan (Being_employed)
- Anders dan planeten hebben planetoden geen bolvorm.

Most occurrences of *hebben* are annotated with the Possession frame (although with a lot of confusion). The Possession frame seems to focuss on material possessions (An Owner has (or lacks) a Possession.) according to its definitons and its main examples. On the other hand many of the Berkeley FN examples also do not focus on real possessions:

- a school has good programmes
- a school has students
- having power
- having ideas

We could just say *hebben* is POSSESSION (great for IA agreement, but not very informative). Or just go on and collect the cases afterwards and decide what we do. Other studies do not report this problem (why not).

Multiword expressions

Separable verbs

Support verbs

References

- [Ahlberg et al., 2014] Ahlberg, M., Borin, L., Dannélls, D., Forsberg, M., Gronostaj, M. T., Heppin, K. F., Johansson, R., Kokkinakis, D., Olsson, L.-J., and Uppström, J. (2014). Swedish framenet++ the beginning of the end and the end of the beginning.
- [Borin and xx, 2010] Borin, L. and xx (2010). Swedish framenet++.
- [Burchardt et al., 2009] Burchardt, A., Erk, K., Frank, A., Kowalski, A., Pado, S., and Pinkal, M. (2009). Using framenet for the semantic analysis of german: Annotation, representation, and automation.
- [Djemaa et al., 2016] Djemaa, M., Candito, M., Muller, P., and Vieu, L. (2016). Corpus annotation within the french framenet: a domain-by-domain methodology. In *Proceedings of LREC 2016*.
- [Hayoun and Elhadad, 2016] Hayoun, A. and Elhadad, M. (2016). The hebrew framenet project the hebrew framenet project. In *Proceedings of LREC 2016*.
- [Subirats, 2009] Subirats, C. (2009). *Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon*, pages 135–162. Mouton de Gruyter, Berlin/New York.

8 Appendix I: agreement per file

Table 7: file agreement (part 1)

theme/genre	filename	nrframes	nr_agreements	nr_of disagreement	agreement rates
dpc dpc-bmm-001087-nl-sen-fn.txt.xml	background-news	59	32	27	0.54
dpc dpc-bmm-001088-nl-sen-fn.txt.xml	background-news	36	18	18	0.50
dpc dpc-bmm-001089-nl-sen-fn.txt.xml	background-news	15	6	9	0.40
dpc dpc-cam-001280-nl-sen-fn.txt.xml	financial	50	22	28	0.44
dpc dpc-eli-000937-nl-sen-fn.txt.xml	financial	54	29	25	0.54
dpc dpc-eli-000940-nl-sen-fn.txt.xml	financial	75	41	34	0.55
dpc dpc-eli-000941-nl-sen-fn.txt.xml	financial	59	28	31	0.47
dpc dpc-eup-000015-nl-sen-fn.txt.xml	financial	143	84	59	0.59
dpc dpc-eup-000017-nl-sen-fn.txt.xml	financial	45	27	18	0.60
dpc dpc-eup-000018-nl-sen-fn.txt.xml	financial	179	90	89	0.50
dpc dpc-eup-000019-nl-sen-fn.txt.xml	financial	37	19	18	0.51
dpc dpc-ind-001632-nl-sen-fn.txt.xml	news	51	30	21	0.59
dpc dpc-ind-001633-nl-sen-fn.txt.xml	news	109	75	34	0.69
dpc dpc-ind-001646-nl-sen-fn.txt.xml	news	105	58	47	0.55
dpc dpc-ind-001649-nl-sen-fn.txt.xml	news	111	66	45	0.59
dpc dpc-ind-001650-nl-sen-fn.txt.xml	news	123	63	60	0.51
dpc dpc-kok-001320-nl-sen-fn.txt.xml	financial	34	14	20	0.41
dpc dpc-kok-001321-nl-sen-fn.txt.xml	financial	28	20	8	0.71
dpc dpc-kok-001324-nl-sen-fn.txt.xml	financial	44	18	26	0.41
dpc dpc-qty-000933-nl-sen-fn.txt.xml	medical	88	48	40	0.55
wiki wiki-1058-fn.txt.xml	wiki	21	9	12	0.43
wiki wiki-11-fn.txt.xml	wiki	87	50	37	0.57
wiki wiki-1181-fn.txt.xml	wiki	44	13	31	0.30
wiki wiki-1250-fn.txt.xml	wiki	7	3	4	0.43
wiki wiki-1327-fn.txt.xml	wiki	13	8	5	0.62
wiki wiki-1400-fn.txt.xml	wiki	80	31	49	0.39
wiki wiki-1541-fn.txt.xml	wiki	64	40	24	0.62
wiki wiki-1550-fn.txt.xml	wiki	27	11	16	0.41
wiki wiki-1553-fn.txt.xml	wiki	10	3	7	0.30
wiki wiki-1566-fn.txt.xml	wiki	44	25	19	0.57
wiki wiki-1609-fn.txt.xml	wiki	19	8	11	0.42
wiki wiki-1617-fn.txt.xml	wiki	32	16	16	0.50
wiki wiki-1625-fn.txt.xml	wiki	9	6	3	0.67
wiki wiki-1635-fn.txt.xml	wiki	47	17	30	0.36
wiki wiki-1652-fn.txt.xml	wiki	19	9	10	0.47
wiki wiki-1799-fn.txt.xml	wiki	51	28	23	0.55
wiki wiki-1808-fn.txt.xml	wiki	58	36	22	0.62
wiki wiki-1820-fn.txt.xml	wiki	7	5	2	0.71
wiki wiki-1823-fn.txt.xml	wiki	142	74	68	0.52
wiki wiki-1846-fn.txt.xml	wiki	43	17	26	0.40
wiki wiki-1941-fn.txt.xml	wiki	90	49	41	0.54
wiki wiki-295-fn.txt.xml	wiki	86	40	46	0.47
wiki wiki-334-fn.txt.xml	wiki	32	18	14	0.56
wiki wiki-342-fn.txt.xml	wiki	9	5	4	0.56
wiki wiki-349-fn.txt.xml	wiki	42	16	26	0.38
wiki wiki-3668-fn.txt.xml	wiki	18	7	11	0.39
wiki wiki-572-fn.txt.xml	wiki	14	9	5	0.64
wiki wiki-60-fn.txt.xml	wiki	6	3	3	0.50
wiki wiki-62-fn.txt.xml	wiki	11	4	7	0.36
wiki wiki-8628-fn.txt.xml	wiki	110	53	57	0.48
wiki wiki-8783-fn.txt.xml	wiki	8	6	2	0.75
wiki wiki-89-fn.txt.xml	wiki	16	6	10	0.38
wiki wiki-90-fn.txt.xml	wiki	19	5	14	0.26
wiki wiki-9809-fn.txt.xml	wiki	10	7	3	0.70

Table 8: file agreement (part 2)

theme/genre	filename	nrframes	nr_agreements	nr_of disagreement	agreement rates
wr WR-P-E-E-0000000018-fn.txt.xml	newsletter	56	28	28	0.50
wr WR-P-E-E-0000000019-fn.txt.xml	newsletter	31	19	12	0.61
wr WR-P-E-E-0000000020-fn.txt.xml	newsletter	24	12	12	0.50
wr WR-P-E-H-0000000009-fn.txt.xml	teletext	48	34	14	0.71
wr WR-P-E-H-0000000013-fn.txt.xml	teletext	77	52	25	0.68
wr WR-P-E-H-0000000036-fn.txt.	44	27	17	0.61	
wr WR-P-E-I-0000000001-fn.txt.xml	websites	49	26	23	0.53
wr WR-P-P-H-0000000004-fn.txt.xml	periodicals	4	3	1	0.75
wr WR-P-P-H-0000000005-fn.txt.xml	periodicals	25	15	10	0.60
wr WR-P-P-H-0000000006-fn.txt.xml	periodicals	22	11	11	0.50
wr WR-P-P-H-0000000007-fn.txt.xml	periodicals	7	5	2	0.71
wr WR-P-P-H-0000000009-fn.txt.xml	periodicals	35	19	16	0.54
wr WR-P-P-H-0000000010-fn.txt.xml	periodicals	15	8	7	0.53
wr WR-P-P-H-0000000011-fn.txt.xml	periodicals	4	0	4	0.00
wr WR-P-P-H-0000000012-fn.txt.xml	periodicals	41	14	27	0.34
wr WR-P-P-H-0000000013-fn.txt.xml	periodicals	5	2	3	0.40
wr WR-P-P-H-0000000014-fn.txt.xml	periodicals	38	21	17	0.55
wr WR-P-P-H-0000000015-fn.txt.xml	periodicals	8	6	2	0.75
wr WR-P-P-H-0000000016-fn.txt.xml	periodicals	3	1	2	0.33
wr WR-P-P-H-0000000017-fn.txt.xml	periodicals	3	1	2	0.33
wr WR-P-P-H-0000000018-fn.txt.xml	periodicals	3	0	3	0.00
wr WR-P-P-H-0000000019-fn.txt.xml	periodicals	61	33	28	0.54
wr WR-P-P-H-0000000021-fn.txt.xml	periodicals	16	10	6	0.62
wr WR-P-P-H-0000000038-fn.txt.xml	periodicals	12	8	4	0.67
wr WR-P-P-H-0000000039-fn.txt.xml	periodicals	60	29	31	0.48
wr WR-P-P-H-0000000040-fn.txt.xml	periodicals	10	6	4	0.60
wr WR-P-P-H-0000000041-fn.txt.xml	periodicals	53	16	37	0.30
wr WR-P-P-H-0000000042-fn.txt.xml	periodicals	9	5	4	0.56
wr WR-P-P-H-0000000043-fn.txt.xml	periodicals	7	7	0	1.00
wr WR-P-P-H-0000000044-fn.txt.xml	periodicals	46	33	13	0.72
wr WR-P-P-H-0000000045-fn.txt.xml	periodicals	11	8	3	0.73
wr WR-P-P-H-0000000046-fn.txt.xml	periodicals	35	16	19	0.46
wr WR-P-P-H-0000000047-fn.txt.xml	periodicals	30	7	23	0.23
wr WR-P-P-H-0000000090-fn.txt.xml	periodicals	12	5	7	0.42
wr WR-P-P-H-0000000091-fn.txt.xml	periodicals	4	4	0	1.00
wr WR-P-P-H-0000000092-fn.txt.xml	periodicals	35	20	15	0.57
wr WR-P-P-H-0000000093-fn.txt.xml	periodicals	8	3	5	0.38
wr WR-P-P-H-0000000094-fn.txt.xml	periodicals	37	20	17	0.54
wr WR-P-P-H-0000000095-fn.txt.xml	periodicals	103	64	39	0.62
wr WR-P-P-H-0000000096-fn.txt.xml	periodicals	16	10	6	0.62
wr WR-P-P-H-0000000097-fn.txt.xml	periodicals	4	3	1	0.75
wr WR-P-P-H-0000000098-fn.txt.xml	periodicals	17	10	7	0.59
wr WR-P-P-H-0000000099-fn.txt.xml	periodicals	17	9	8	0.53
wr WR-P-P-H-0000000100-fn.txt.xml	policy	5	3	2	0.60
wr WR-P-P-I-0000000025-fn.txt.xml	policy	107	56	51	0.52
wr WR-P-P-I-0000000026-fn.txt.xml	policy	14	8	6	0.57
wr WR-P-P-I-0000000027-fn.txt.xml	policy	5	3	2	0.60
wr WR-P-P-I-0000000028-fn.txt.xml	policy	5	3	2	0.60
wr WR-P-P-I-0000000029-fn.txt.xml	policy	9	2	7	0.22
wr WR-P-P-I-0000000030-fn.txt.xml	policy	4	0	4	0.00
wr WR-P-P-I-0000000031-fn.txt.xml	policy	6	3	3	0.50
wr WR-P-P-I-0000000032-fn.txt.xml	policy	11	5	6	0.45
wr WR-P-P-I-0000000033-fn.txt.xml	policy	99	48	51	0.48
wr WR-P-P-I-0000000034-fn.txt.xml	policy	18	7	11	0.39
wr WR-P-P-I-0000000035-fn.txt.xml	policy	23	10	13	0.43
wr WR-P-P-I-0000000235-fn.txt.xml	policy	51	28	23	0.55
ws WS-U-E-A-0000000001-fn.txt.xml	financial	12 110	74	36	0.67
ws WS-U-E-A-0000000007-fn.txt.xml	financial	160	101	59	0.63
ws WS-U-E-A-0000000009-fn.txt.xml	financial	204	120	84	0.59
ws WS-U-E-A-0000000010-fn.txt.xml	financial	174	104	70	0.60
ws WS-U-E-A-0000000012-fn.txt.xml	financial	174	99	75	0.57
ws WS-U-E-A-0000000017-fn.txt.xml	financial	186	122	64	0.66