

Levels of organisation in ontology verbalisation

Sandra Williams, Allan Third and Richard Power

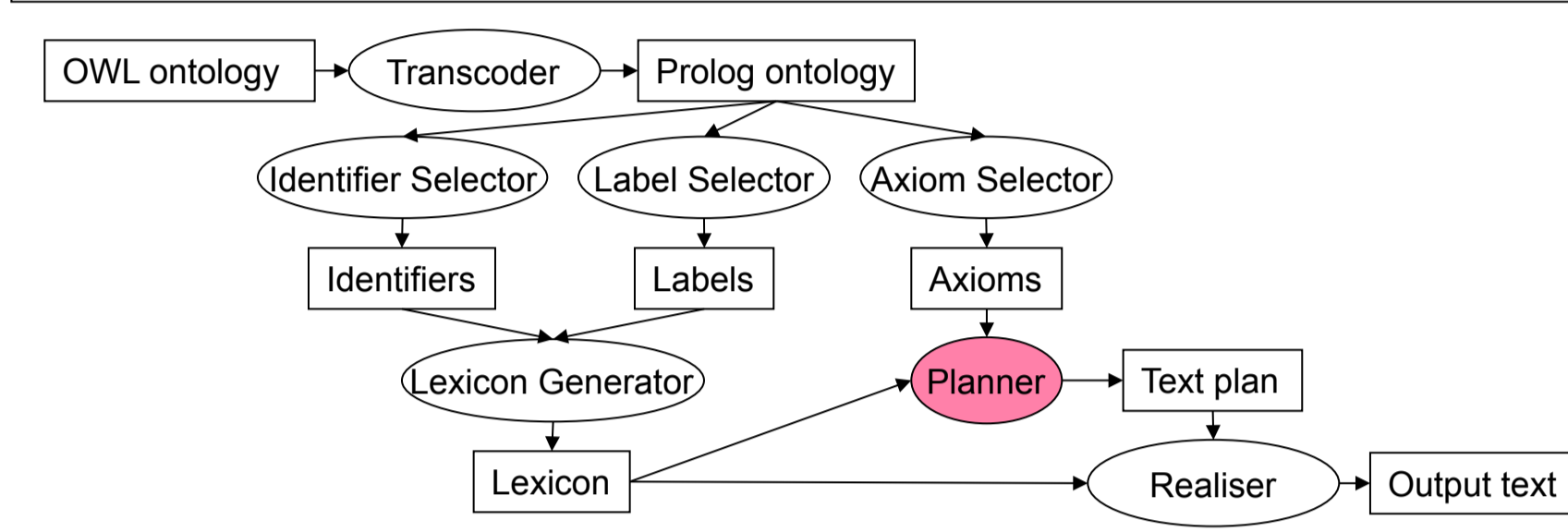


Overview

The planning component of the SWAT Tools system generates text from Web Ontology Language (OWL), a process known as verbalisation (see the architecture diagram below). It organises text at a number of levels inspired by encyclopedia entries:

- alphabetical arrangement of headings for entries
- organisation of sub-headings according to logical content
- aggregation of sentences that have parts in common
- hyperlinking of phrases to entry headings.

A consequence of text organisation is *repetition* of statements that are relevant to more than one entry; this means that an organised text is much longer than one in which statements are simply listed. The trade-off between organisation and brevity is investigated in a task-based user study.



1. Overview

Higher Levels of Organisation

Top Level

- Alphabetically ordered entry headings extracted from SWAT Tools lexical entries

• Data for entries is selected from the ontology if the class, property or individual of the heading occurs as a top-level argument of a statement.

2nd Level

- Subheadings:
 - Definition
 - Typology
 - Description
 - Distinctions
 - Examples

SETA APPENDAGE CEPHALOTHORAX (class)

Typology
A seta appendage cephalothorax is a seta.

Description
A seta appendage cephalothorax is part of an appendages cephalothorax.

Examples
The following are seta appendage cephalothorax: male palpal femoral thorns, female palp femoral thorns and spd 0000203a, and so on (5 items in total).

SETA CARAPACE (class)

Lower Levels of Organisation

3rd Level

- Aggregated statements with identical logical structures.

- Truncation of long lists reduces length somewhat.

4th Level

- Hyperlinked phrases.

2. Higher Levels of Organisation

3. Lower Levels of Organisation

SWAT Natural Language Tools

Analyse your OWL ontology, build a lexicon from it, or convert it into English sentences or definition paragraphs.

Select the ontology file

Select the output format

- Alphabetical English glossary (class, individual and property definitions)
- English sentences (one sentence per OWL axiom)
- Prolog terms (translate OWL to Prolog)
- Lexicon (lexical entries for class, individual and property names)
- OWL patterns (frequency counts)
- OWL/XML, with verbalisations as "SWATDescription" annotations.

4. SWAT NL Tools: swat.open.ac.uk/tools

A maAmp base is part of a maAmp.
A spd 0000161 is a chemosensory seta.
A spd 0000161 is part of a cymbium.
A cymbium spical setae is a seta appendage cephalothorax.
A cymbium spical setae is part of a cymbium.
An alveolus is part of a cymbium.
An alveolus retromargin is part of an alveolus.
An alveolus promargin is part of an alveolus.
A spd 0000166 is part of a cymbium.
A cymbial dorsal processes is part of a cymbium.
A copulatory bulb is part of a cymbium.
A petiolus is part of a copulatory bulb.
A basal hematodocha is part of a copulatory bulb.
A subtegulum is part of a copulatory bulb.
A median apophysis is part of a copulatory bulb.
A tegulum is part of a copulatory bulb.
A median hematodocha is part of a copulatory bulb.
An embolic division is part of a copulatory bulb.
An embolus is part of an embolic division.
A spermaphor is part of a copulatory bulb.
A median apophysis is part of a copulatory bulb.
A conductor is part of a copulatory bulb.
A terminal hematodocha is part of a copulatory bulb.
A sclerite unspecified is part of a copulatory bulb.
A maAmp shaft is a spigot shaft.
A maAmp shaft is part of a maAmp.
A PI base is a spigot base.
A PI base is part of a PI.
A cheliceral boss is part of a paturon.
A promargin is part of a paturon.
A retromargin is part of a paturon.
A cheliceral gland is part of a paturon.
A stitulatory files is part of a paturon.
A PI shaft is a spigot shaft.
A PI shaft is part of a PI.
A PLS Ac shaft is a spigot shaft.
A PLS Ac shaft is part of a PI.

5. Unorganised Output, 4,803 words (9 pages of A4)

SETA APPENDAGE CEPHALOTHORAX (class)

Typology
A seta appendage cephalothorax is a seta.

Description
A seta appendage cephalothorax is part of an appendages cephalothorax.

Examples
The following are seta appendage cephalothorax: male palpal femoral thorns, female palp femoral thorns and spd 0000203a, and so on (5 items in total).

SETA CARAPACE (class)

Typology
A seta carapace is part of a cuticle carapace.

Description
Scale carapaces, and hair carapaces are seta carapaces.

6. Organised Output, 7,746 words (25 pages of A4)

Research Question and Hypotheses

Research Question

Does organisation help people understand and navigate a text in spite of the increased length?

Hypotheses

Navigation

- People will judge information easier to find in an organised text.

Comprehension

- People will answer questions more accurately with an organised text

7. Why organise a text?

Experiment

Task

Navigate a text generated by SWAT NLP tools to locate information and answer five questions.

Judge how difficult the information was to find.

Materials

Organised output generated from spider anatomy ontology (as above).
Control: unorganised output from same ontology (see above).

Participants

Group 1: 29 people from SIGGEN and SIGdial used the organised text.
Group 2: 28 people from SIGGEN and SIGdial used the unorganised text.
None had expert knowledge of spider anatomy.

Questions

1. What is a tarsus?
2. Name 3 kinds of spigot shaft.
3. What is a palp?
4. Name 2 kinds of silk cable.
5. How many kinds of seta appendage cephalothorax are there in total?

8. Task-based Experiment

INSTRUCTIONS

- Questions in the survey relate to sentences on a separate Web page.
- When given the link to the Web page, right-click it and select 'Open Link in New Window'. Keep it open throughout the survey so that you can answer the remaining questions.
- Search for information using any method you like.
- The Web page is long, so you may need to scroll down to find information.
- Answer questions with information from the Web page. Please do this EVEN IF YOU KNOW A DIFFERENT ANSWER (we want to know what you actually found on the Web page rather than how much you know).
- Following each question, you will be asked to judge how easy it was to find the information.

★ 1. What is a tarsus?

A tarsus is part of a leg.

★ How hard was it to find the information?

Very Easy

Easy

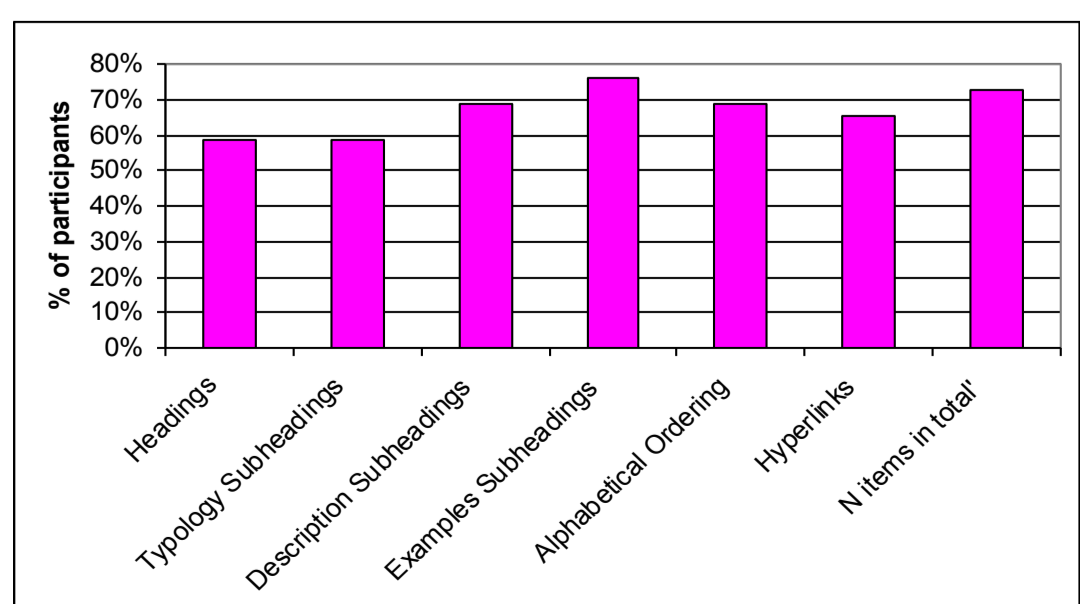
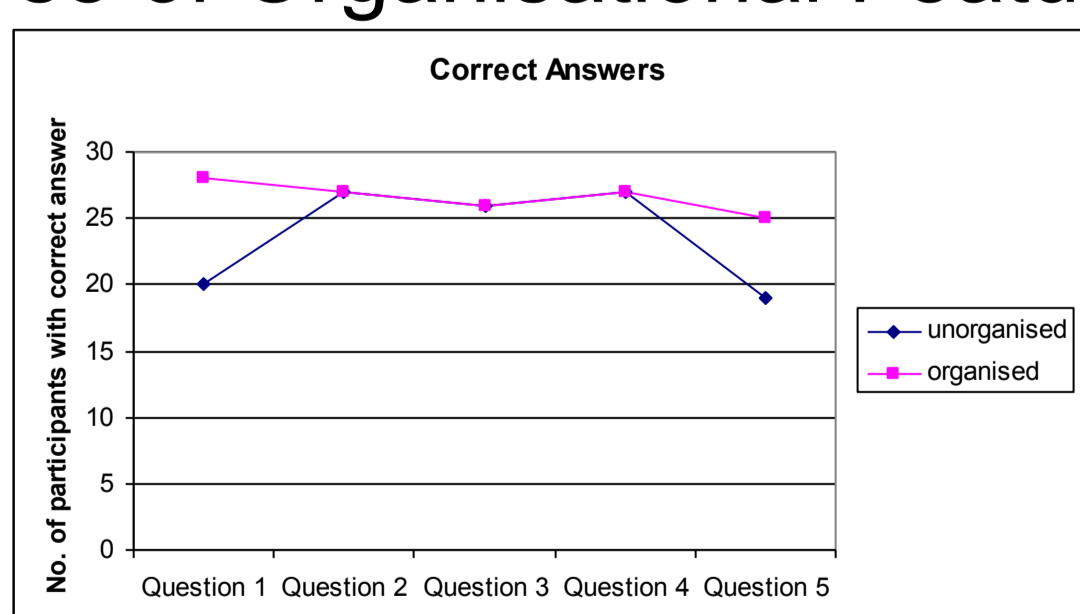
Neither hard or easy

Hard

Very Hard

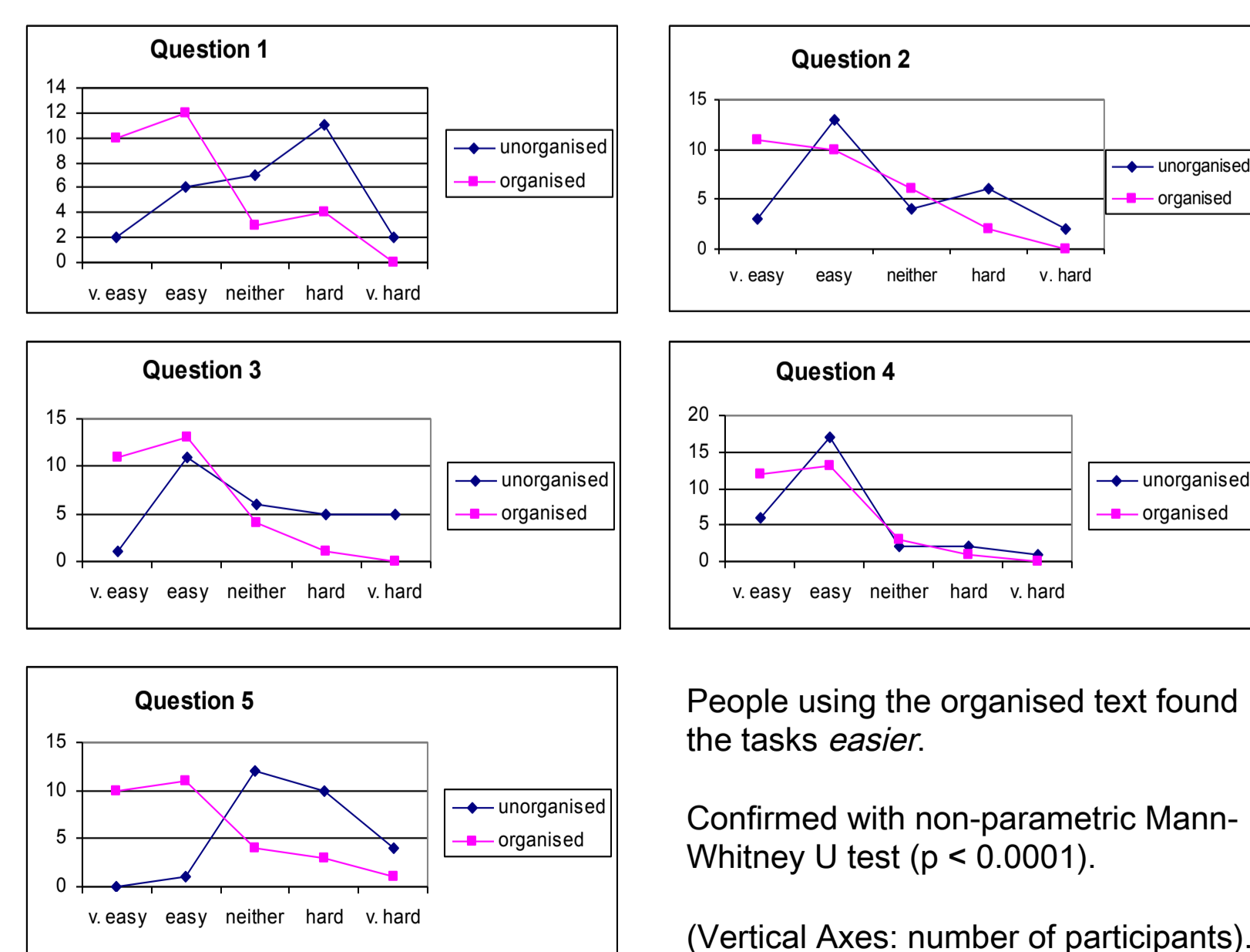
9. The Survey

Correct Answers and Use of Organisational Features



10. Results for Correct Answers / Organisational Features

Difficulty Judgements



11. Results for Difficulty Judgements

Discussion and Conclusion

Our earlier studies with bioinformatics experts indicated that they preferred *natural* English over strict fidelity to OWL semantics.

Headings, sub-headings and aggregation introduce:

- a further move towards naturalness that was not directly encoded into OWL
- repetitions that render the organised text much longer.

We found no evidence that people perform a navigation task more accurately with an organised text (comprehension hypothesis was *not* supported).

However, people viewing the organised text found the tasks easier (navigation hypothesis was supported).

Perhaps people do whatever is necessary to achieve a desired level of performance, so that when provided with superior tools they achieve the same result but the perceived effort is less.

All organisational features were considered useful. None stood out as particularly important.

We assume people prefer a verbalisation that is organised like a naturally occurring text of the appropriate genre.

The organised text is longer but there is no loss of performance.

12. Discussion / Conclusion