

Ontology-Grounded Structured Prediction for Dental CBCT Reporting

Luca Lumetti · Mattia Di Bartolomeo · Arrigo Pellacani · Alexandre Anesi
Costantino Grana · Federico Bolelli
University of Modena and Reggio Emilia, Italy



Why structured reporting?

- Free-text wording varies between clinicians, so the same finding can be written in many ways.
- Public CBCT data such as **ToothFairy3** (532 volumes, 77 classes) targets segmentation, not reporting.
- Goal: predict every finding as a **verifiable instance** of type, location and qualifiers.

Idea: a report is a set of ontology-grounded findings.

A clinician-grounded resource

Core contribution

529

CBCT volumes

893

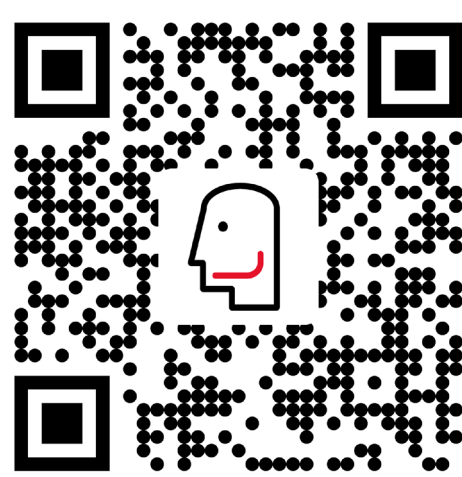
clinical reports

5,703

findings

13

finding types



Released as **ToothFairy4**, the new version of the ToothFairy dataset.
2 radiologists + 6 maxillofacial surgeons
Clinician-designed OWL/RDF ontology · SHACL validation · Clinician-reviewed labels



The ontology: many wordings, one checkable finding

Core contribution

1 FREE TEXT

REPORT A

"Mandibular canal with a predominantly lingual course, close to teeth 37 and 38."

REPORT B

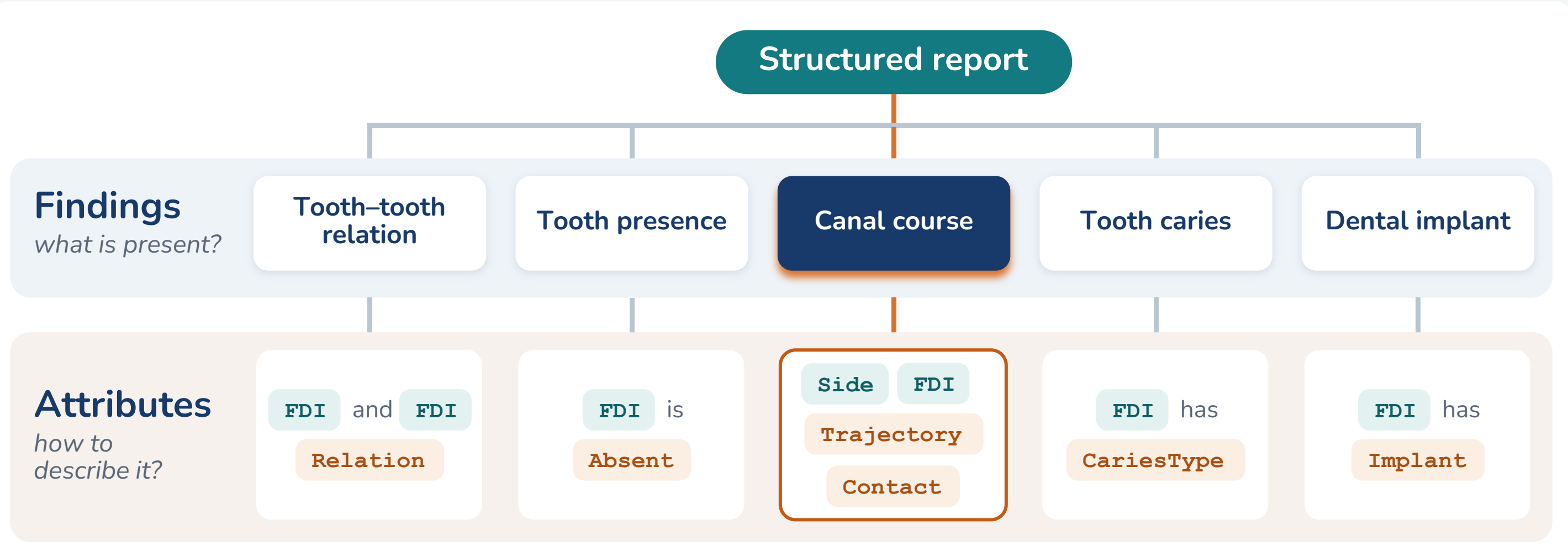
"Left IAN canal runs lingually, in contact with the apices of 37–38."

REPORT C

"Lingual course of the left mandibular canal; roots of 3.7 and 3.8 touch it."

Three wordings, one clinical finding.

2 ONTOLOGY · OWL/RDF + SHACL



Each finding type defines its own attributes and allowed values; SHACL shapes reject invalid combinations.

3 STRUCTURED FINDING

FINDING

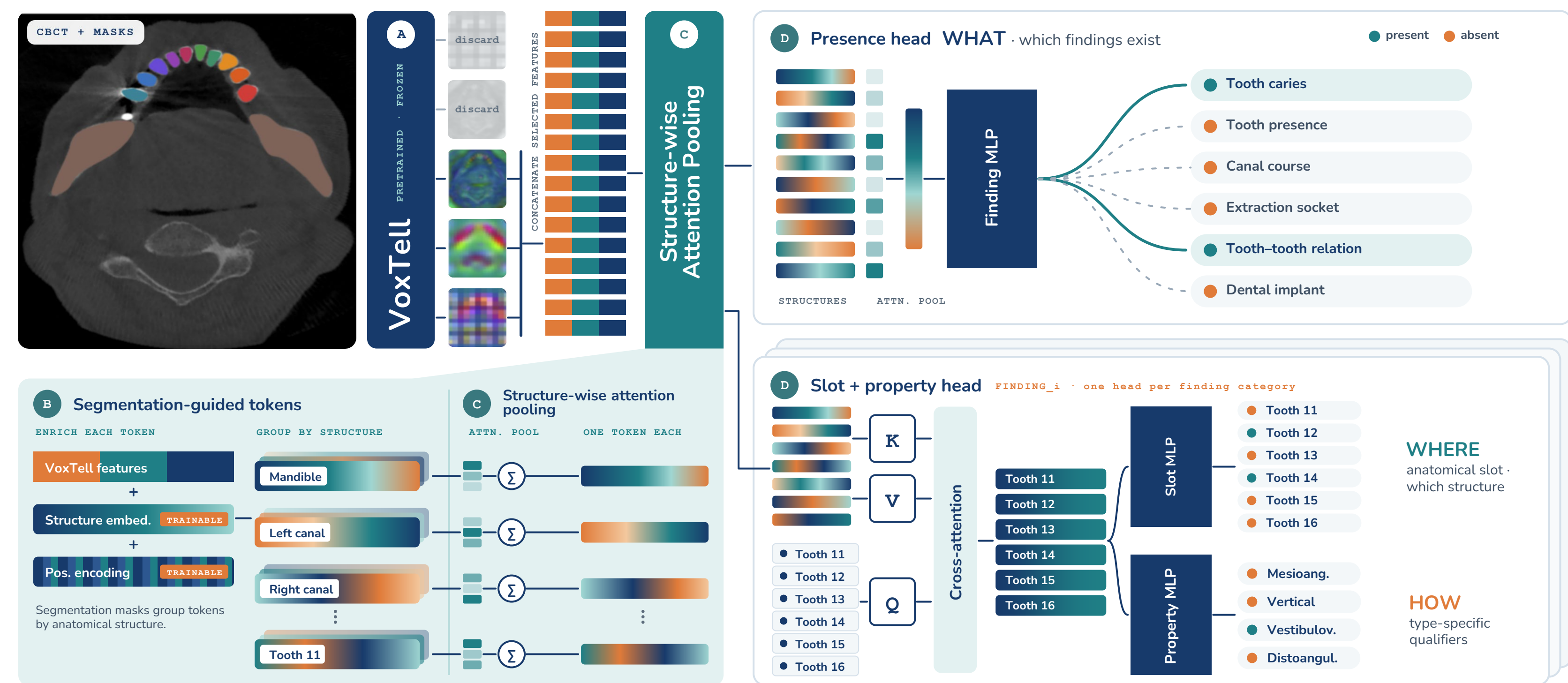
Canal course

ATTRIBUTES

Side Left
FDI 37, 38
Trajectory Prevalently lingual
Contact Intimate

Same output for A, B and C.

A structure-aware prediction pipeline



A

Frozen VoxTell features

Pretrained, frozen CBCT encoder; features from levels 3, 4 and 5 are fused across scales.

B

Segmentation-guided tokens

Segmentation masks group voxel features by anatomical structure, adding structure and position embeddings.

C

Structure-wise attention pooling

Attention pools each structure into one embedding, focused on its most informative voxels.

D

SHACL-defined heads

Heads predict finding presence, anatomical slots and properties, so every output is a valid ontology instance.

What the results show

5-fold CV means · fixed held-out test set

Metric	What it checks	nnU-Net	GT masks	Gap to GT
pres-F1	Finding detected	71.2	72.0	−0.8
inst-F1	+ correct location	43.3	46.7	−3.4
char-acc	+ correct qualifiers	51.0	85.4	−34.4
rep-F1	Whole per-patient report	49.8	53.8	−4.0
slot-F1	Location only	51.7	53.4	−1.7
prop-F1	Qualifiers only	50.1	61.2	−11.1

Highlighted rows: gap > 5 points, where segmentation quality matters most.

DETECTION · pres-F1

71.2 vs 72.0

Barely changes with predicted masks.

CHARACTERIZATION · char-acc

51.0 vs 85.4

Sensitive to segmentation quality.

With perfect masks, detection caps the pipeline; with predicted masks, segmentation does.



Yggdrasil

The annotation platform behind the dataset

Yggdrasil is the AlmageLab medical-imaging platform used to collect, view and annotate the **ToothFairy4** data, linking dental imaging with clinician-authored reports.
A read-only live demo is open to everyone.

