

t - tile  
 td - tile, description  
 tdc - tile, description, color  
 tds - tile, description, shape  
 tdf - tile, description field (where a field is any spatial description)  
 tdo - tile, description, other (length, how it is turned, etc)  
 tp - tile, pronoun  
 l - landmark  
 e - words that refer back to a previous episode  
 ld - landmark description  
 lf - landmark field  
 r - relation word (the target is the head)  
 - - direction of relation (when the landmark is described as relative to the target)  
 o - other linguistic material  
 q - words uttered after the beep, or abandoned words (e.g. grün- gelben)  
 n - words that fall in a negative scope (e.g., nicht das rote)  
 x - other non-linguistic material (shouldn't happen, but just in case)

example:

äh das objekt ist ein rotes kreuz ganz links neben dem blauen t  
 o t t t tdc tds tdf tdf r l ldc lds

what if there are multiple landmarks? Then we need to number them, along with their relational word to the target so we know which landmark words belonged to which relational word.

example:

das rote kreuz ist neben einem blauen t und liegt unter einem grünen objekt  
 t tdc tds t r1 l1 l1dc l1ds t t r2 l2 l2dc l2ds

what about relational words that are actually more than one word? Each word is annotated with an r.

example:

das rote kreuz links von einem blauen T  
 t tdc tds r r l ldc lds

what if the landmark is described relative to the target? then the relative word must be marked with a -

example:

neben das ziel objekt liegt einem blauen T  
 r- t t t l l ldc lds

example:

As we are using ASR output, we are annotating words that are possibly wrong. We will also mark words as correct or incorrect in a separate column, to give us an artificial WER. The tags described above should be given to the word based on the ground truth, i.e., what was actually said and not what the recognized word might mean. E.g., if the person said grün, but größer was recognized, it is tagged as tdc.

what if someone references an object in a previous episode? Then those words should be marked with e.

example:

über den stand der gerade eben gesucht wurde befindet sich grauen holzblock

r   le   le   le   le   le   le   le   t   t   t   t

t <- I is normal

I <- t is -

Notes:

Interesting that the way someone does the references is quickly grounded, so the wizard can easily get ones wrong if he doesn't like the way things are going.

Generally, they use colors, shapes, and global spatial placements with the occasional landmark. All four of those strategies seem to be uttered in random order.

repairs:

r6.29.p1, r6.30.p1, r1.17.p2, r3.24.p2, r3.31.p2, r3.38.p2, r6.21.p2 (annotated for actual objects)

r6 would sometimes say where both tiles were, e.g., beiden befinden sich in der oberen ecke

r7.57.p2 - the reference object was really far away

r8 was terrible. Way too many words. ASR did not perform well.