

## A Examples of HAN Attention

The examples were taken from the Spanish-English TED talks corpus. We show the behavior of the attention function of HAN. First, we show the attention to context for HAN encoder and HAN decoder respectively. Second, we show the multi-head attention only for HAN decoder (English) for better understanding.

### A.1 Encoder and Decoder Attention

Currently Translated Sentence	
Src.:	y <b>toqué</b> el primer movimiento del concierto para violín de Beethoven .
Ref.:	and I <b>played</b> the first movement of the Beethoven Violin Concerto .
Base:	and I <b>touched</b> the first move from the concert to Beethoven .
Cache:	and I <b>touched</b> the first move of Beethoven 's violin .
HAN:	and I <b>played</b> the first move of Beethoven 's violin .

  

Context from Previous Sentences	
HAN decoder context with target. <i>Query: played</i> (En)	
s <sup>t-3</sup>	and he was talking about invisible demons and smoke , and how someone was sleeping with him .
s <sup>t-2</sup>	and I felt fear , not for me , but fear that I was going to lose it ...
s <sup>t-1</sup>	so I just started playing .
HAN encoder context with source. <i>Query: toqué</i> (Es)	
s <sup>t-3</sup>	y hablaba de demonios invisibles y humo , y de cómo alguien lo estaba envenenando mientras dormía .
s <sup>t-2</sup>	y yo sentí miedo , no por mí , sino miedo de que iba a perderlo ...
s <sup>t-1</sup>	por ello sólo empecé a tocar .

Table 1: In this example, the HAN model disambiguates correctly the word “toqué”, which can be translated as “touched” or “played”. We can see that the HAN decoder uses the semantically close word “playing” from the previous sentence. In similar manner, the HAN encoder focused on “tocar” which is coherent with “toqué”.

### A.2 Multi-Head Attention

Currently Translated Sentence	
Src.:	y < ellos > estarían tan compenetrados en la partida de dados porque los juegos son tan atractivos ...
Ref:	and <b>they</b> would be so immersed in playing the dice games because games are so engaging ..
Base:	and <b>you</b> would be so < unk > in the start of it because games are so attractive ...
HAN:	and <b>they</b> would be so < unk > in the start of dice because games are so attractive ...

  

Context from Previous Sentences. <i>Query: they</i>	
Head 2: Attention to the antecedent “people” in $s^{t-3}$ .	
s <sup>t-3</sup>	people suffered . people suffered .
s <sup>t-2</sup>	it was an extreme situation . they needed an extreme solution .
s <sup>t-1</sup>	so , according to Indyk , the games of dice and a policy was established throughout the kingdom : one day , everybody would eat , and the next day , everybody would eat .
Head 4: Attention to the same pronoun “they” in $s^{t-2}$	
s <sup>t-3</sup>	people suffered . people suffered .
s <sup>t-2</sup>	it was an extreme situation . they needed an extreme solution .
s <sup>t-1</sup>	so , according to Indyk , the games of dice and a policy was established throughout the kingdom : one day , everybody would eat , and the next day , everybody would eat .
Head 7: Attention to verbs that conjugate with “they”	
s <sup>t-3</sup>	people suffered . people suffered .
s <sup>t-2</sup>	it was an extreme situation . they needed an extreme solution .
s <sup>t-1</sup>	so , according to Indyk , the games of dice and a policy was established throughout the kingdom : one day , everybody would eat , and the next day , everybody would eat .

Table 2: This example displays the translation of Spanish pronoun “ellos”, which is a dropped-pronoun which is implicit in the verb conjugation of “estarían”. As we can observe, HAN translates correctly the dropped-pronoun into the English “they”. Each head focuses on a different aspect during translation, for example head 2 seems to attend to the antecedent of the pronoun “people” in the third previous sentence, head 4 attends to the same pronoun on the second previous sentence, and head 7 attends to different verbs on all previous sentences.

Currently Translated Sentence

Src:	y como resultado <b>construimos</b> relaciones sociales más fuertes .
Ref:	and we actually <b>build</b> stronger social relationships as a result .
Base:	and as a result , we <b>construct</b> stronger social relationships .
HAN:	and as a result , we <b>build</b> stronger social relationships .

Context from Previous Sentences. *Query: build*

Head 1: Attention to related words “ <i>construimos</i> ”, “ <i>trust</i> ”...	
s <sup>t-3</sup>	and the reason is that it demands a lot of trust to play a game with someone .
s <sup>t-2</sup>	we trust that they 're going to spend their time with us that they 're going to play under the same rules as the same goal , they 're going to stay in the game all the way down .
s <sup>t-1</sup>	so playing a game together actually builds ties and trust and cooperation .
Head 4: Attention to a similar translation “ <i>builds</i> ” in $s^{t-1}$	
s <sup>t-3</sup>	and the reason is that it demands a lot of trust to play a game with someone .
s <sup>t-2</sup>	we trust that they 're going to spend their time with us that they 're going to play under the same rules as the same goal , they 're going to stay in the game all the way down .
s <sup>t-1</sup>	so playing a game together actually builds ties and trust and cooperation .

Table 3: This example displays the translation of the ambiguous Spanish word “*construimos*”, which can be translated as “*construct*” or “*build*”. HAN translates this word correctly according to the context using for example related words “*trust*”, “*ties*”, and “*cooperation*” on previous sentences with *head 1*, and a previous translation “*builds*” in the previous sentence with *head 4*.

Currently Translated Sentence

Src:	antes de los fantásticos controladores de <b>juegos</b> teníamos tabas de oveja .
Ref:	before we had awesome <b>game</b> controllers , we had sheep 's knuckles .
Base:	before the fantastic <i>TV</i> controllers , we had < unk > .
HAN:	before the fantastic <b>game</b> controllers , we had < unk > .

Context from Previous Sentences. *Query: game*

Head 3: Attention to similar word “ <i>game</i> ” in $s^{t-3}$	
s <sup>t-3</sup>	we have to begin to make the real world more like a game .
s <sup>t-2</sup>	I was inspired by something that happened 2,500 years ago .
s <sup>t-1</sup>	these are ancient dice , made out of sheep UNK . right ?
Head 5: Attention to a related word “ <i>dice</i> ” in $s^{t-1}$	
s <sup>t-3</sup>	we have to begin to make the real world more like a game .
s <sup>t-2</sup>	I was inspired by something that happened 2,500 years ago .
s <sup>t-1</sup>	these are ancient dice , made out of sheep UNK . right ?

Table 4: This example shows the translation of the Spanish word “*juegos*”. The baseline translates it incorrectly, while HAN translates it correctly by spotting a similar translation “*game*” in the third previous sentence with *head 3*, and a related word “*dice*” on previous sentence with *head 5*.