Комментарии к токенизации.

1. Слова этот, тот, эти, те (this, that, these, those) можно токенизировать и как demonstrative pronouns и как demonstrative determiners. Они отражают положение предмета относительно нашего расположения в пространстве.

Если мы примем эти слова за местоимения (pronouns), тогда эти слова выполняют функции существительного, когда мы не хотим называть само существительное (в т.ч. для того, чтобы сократить длину предложений или когда очевидно, о каком существительном мы говорим).

Если мы примем эти слова за детерминативы (determiners), тогда они ставятся перед существительным для того, чтобы ввести, обозначить их.

Примеры:

* demonstrative pronouns – Look at the dress | I want to buy *that.*
* demonstrative determiners - Look at the dress | I want to buy *that* one.

Тогда для аннотаций

- 00:02:33.034 тот – demonstrative pronoun

- 00:02:33.912 этот – determiner

2.) Аналогичные размышления можно привести для possessive pronouns и determiners, слов «мой», «твой» и т.д. Различия будут более ощутимы в английском: притяжательными местоимениями корректнее было бы называть слова «mine», «yours», а словами «my» и «your» - детерминативы. Следовательно, для аннотации

(00:03:09.754; 00:03:13.890) его – determiner

(00:03:52.672; 00:03:58.136) их (дважды) - pronoun

1. Все равно (00:02:30.814) – конструкция, которая может являться:
   1. частицей, используемой для усиления противопоставления ранее сказанному (=все-таки)
   2. наречием, синонимичным конструкции “при любых обстоятельствах, в любом случае”

Мне кажется, конструкцию “все равно” можно рассматривать как частицу, которая показывает, что этот лес действительно не подходит для охоты. Однако так как у говорящего наблюдается слабая связь между репликами (00:02:25.386; 00:02:29.4189) и (00:02:30.626; 00:02:33.272), и первая реплика довольно сумбурна, я отнесла “все равно” к наречиям со значением “при любых обстоятельствах”.

1. Нет (форма нету). Это слово в русском зыке может быть как частицей, так и союзом и предикативом. В аннотации, начинающейся в 00:02:47.408, очевидно, что «нет» выполняет функцию сказуемого, следовательно, часть речи – предикатив. Но в списке возможных частей речи предикатив отсутствуют. Так как большая часть слов, относящихся к предикативам, являются глаголами, обозначим слово «нет» как глагол.
2. Всё. Аннотация (00:03:04.317; 00:03:04.649). В данном случае конструкция «и всё» является маркером конца предложения (законченности мысли говорящего). Это не местоимение, так как за ним нет значения существительного и в этом контексте нельзя предположить, что слово изменяется. Это не союз (так как он не соединяет части сложного предложения) и не частица (так как не усиливает никаких значений в предложении). Тогда, используя метод от противного, я считаю, что это слово – наречие.
3. Вводные слова. Для вводных слов не предусмотрено выбора в слое pos, поэтому необходимо выбрать для них другое обозначение. (информация с dic.academic.ru). Морфологически вводные слова бывают

а) глагольные – скажем, помнится – т.е. те слова, которые изначально были глаголами

б) именные – во-первых, к счастью

в) наречные – короче, действительно

Следовательно, я буду токенизировать вводные слова в соответствие с этой классификацией:

(00:02:11.078; 00:02:11.171) значит – глагол

(00:03:49.038; 00:03:49.464) значит - глагол

(00:03:47.544; 00:03:47.993) например – наречие

(00:04:07.864; 00:04:10.237) кстати - наречие

1. «То что». В аннотации с этими тайм-кодами (00:03:52.672; 00:03:58.136) «то что» по смыслу похоже на «потому что». Поэтому, как мне кажется, эти два слова следует рассматривать как единый токен, соответствующий токену «потому что» - conjunction.