Annotation Guidelines for *The (Living) Dead*

ENG287H1F, FALL 2013

1. GENERAL OVERVIEW

James Joyce's short story "The Dead" is to be annotated using TEI, an XML language. XML encloses a passage (span) of text with start and end markers. A start marker includes the tag name followed by a series of attribute/value pairs with the values in quotation marks. The end marker includes just the tag name, preceded by a forward slash.

```
<tag attribute1="value1" attribute2="value2" …>Some text.</tag>
```

TEI tags may appear embedded within other TEI tags in some circumstances in this text (for example, when a passage of direct discourse occurs within a span of FID). Note, however, that any tags that start within the span of some other tag must finish there—they must, in other words, be properly "nested." Thus, the below is NOT valid XML:

```
<tag1>Some text, <tag2>some more text,</tag1> even more text.</tag2>
```

Rather, the following is valid:

```
<tag1>Some text, <tag2>some more text,</tag2> even more text.</tag1>
```

Our annotation will include two tags, the **said** tag and a **clue** tag. **said** tags can contain **clue** tags and may also contain other **said** tags.

2. THE SAID TAG

A **said** tag should be placed around any span in which voice or thoughts of a character are being reported by direct or free indirect discourse (FID), either by the narrator or by another character. (Spans where the voice or thoughts of a character are summarized or speculated upon by the narrating entity should NOT be surrounded with the **said** tag; it should only be used when a character's actual words are being explicitly reported.) **said** spans in general should be no shorter than a clause (they should break at commas, periods, and occasionally words which separate clauses such as *and* or *that*) and they can be as long as a paragraph. (For multiple consecutive paragraphs, use multiple **said** tags so as to not violate XML nesting rules described above). Narrative elements which indicate speech (but are not actual instances of it) such as "she said" should not generally be included in this tag, unless they are obviously being spoken by a character (i.e. not by the narrator, as is typical).

NB: any words not surrounded by a **said** tag are taken to be the words of the narrator.

The **said** tag has three attributes.

2.1 *Discourse* attribute

The *discourse* attribute is used to express how the words enclosed in the **said** tag are being reported. A "direct" value indicates that they are being reported as direct discourse, meaning that the entity (the narrator

or character) introducing these words is presenting them as the exact words or thoughts of the character they are quoting. The use of first-person pronouns such as "I," the use of the present tense, and adjacent phrases such as "she said" or "he thought" can be indicators of direct discourse. While quotation marks are normally a clear indicator of direct discourse, Joyce does not use them. Instead, he generally uses an em-dash (—) to mark paragraphs that begin with direct discourse. He does not mark the end of passages of direct discourse, nor the beginning of passages of direct discourse that start anywhere other than the beginning of a paragraph. For example:

```
—<said discourse="direct" …>Bring me a plate of peas,</said> he said, <said
discourse="direct" …>and a bottle of ginger beer.</said>
```

The other possible value for the *discourse* attribute is "free", which indicates free indirect discourse (FID). FID is a mixture of narration and direct discourse: like narration, it is in the third person and the past tense; but unlike narration, it contains words and expressions that indicate subjective or personalized aspects clearly distinct from the narrator's style. Otherwise stated, in FID, some aspects of a character's speaking style or emotional state are mixed in with in the narrative voice.

NB: for a passage to qualify as FID, the narrator must do more than simply describe a character's state of mind (e.g. *he was depressed*); rather, a passage worthy of tagging as FID must necessarily contain words that only someone in that state of mind might use (e.g. *he was hating the whole damn stinking dirty world*).

2.1.1 Nesting Said Tags

Said tags may be embedded within one another when a character whose words or thoughts are being reported in turn reports the words or thoughts or another character—for example, when we are taken into a character's thoughts via FID, and in these thoughts the words of another character are reported. One might tag such a passage as follows:

```
<said discourse="free" …>She no longer cared about the unfathomable cruelty of
Robert's calling her a <said discourse="direct" …>"silly fool"</said>.</said>
```

2.2 *Aloud* Attribute

The **aloud** attribute indicates whether or not the expression is actual verbalized speech ("true") or the unspoken internal thoughts of the characters ("false"). Passages in FID are most often silent and passages in direct discourse are most often spoken aloud—but exceptions do occur.

2.3 *Who* attribute

The **who** attribute indicates the character whose speech is being reported in a **said** span. Unique IDs for each character are given at the end of the TEI file. XML syntax requires that a # character be inserted before this ID when assigning a value to the **who** attribute. For example, if the unique ID for a character named Terry were "terry", correct syntax would be as follows:

```
Terry thought, <said who="#terry" …>"I've gone and done it this time"</said>.
```

If the character speaking is not on the list at the end of the TEI file, use "other". If multiple characters are speaking, use "group" (you do not need to list each character in your tagging). If it is not clear from the

context whose words are being quoted or summarized, use "unknown." The narrator is not listed as a character; note again, however, that all words *not enclosed* by **said** tags are taken to belong to the narrator.

3. THE CLUE TAG

A **clue** tag *must* be included within every passage tagged as FID. It is used to indicate the specific words or expressions that lead the annotator to tag the passage as FID. Use the **clue** tag to surround the words and expressions that demonstrate subjective or personalized aspects clearly distinct from the narrator's style.

When the clue consists of an expression (or even multiple consecutive expressions) rather than a single word, the entire span should be tagged, e.g.

```
<said discourse="free" …>He was <clue>hating the whole damn stinking dirty
world</clue></said>
```

But if only one word indicates the situation, tag only that word:

```
<said discourse="free" …>He was cleaning the <clue>damn</clue>toilet
again</said>
```

In general, **said** spans should be larger than **clue** spans; please do not make your entire **said** span a **clue** span simply to satisfy the condition that each span of FID needs one.

4. OTHER NOTES

Be careful not to delete any text while adding annotations. Re-read the TEI after you have put in the tags to make sure it still makes sense, and that all of James Joyce's words appear, unaltered.

There are several instances in "The Dead" where characters sing or recite passages of verse. Because these words do not belong to the characters speaking them, please leave all such passages *un-tagged*.