Annotation Guidelines for *The Brown Stocking*

1. GENERAL OVERVIEW

The selected portion of *To the Lighthouse* is to be annotated using TEI, an XML language. XML encloses a passage (span) of text with start and end markers. A start marker (or tag) includes a *tag* and then a series of attribute/value pairs with the values in quotation marks. The end marker includes just the tag, 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 2 tags, the **said** tag and a **clue** tag. **Said** tags will contain **clue** tags and may also contain other **said** tags.

2. THE SAID TAG

A **said** tag should be placed around any span where the voice or thoughts of a character are being reported, either by the narrator or by another character. (Spans where the voice or thoughts of a character are merely being *speculated* upon should NOT be surrounded with the **said** tag; it should only be used when a character's 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 the narrator, as is typical). Note that anything not included in a **said** tag is taken to be the voice 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. Obvious examples of direct discourse will be signalled by quotation marks, though not all

direct discourse will be placed in quotation marks. 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.

An "indirect" value indicates that the words enclosed in the **said** tag are being reported as indirect discourse, meaning that the span presents the words or thoughts of a character rewritten in the words of the entity (the narrator or another character) who introduces these words. In a third person narrative such as this, indirect discourse is likely to take one of two forms: either the narrator paraphrasing the thoughts of a character in his or her own voice, or one of the characters reporting the speech or thoughts of another character in his or her own words. In the latter case, the <said discourse="indirect"…> tag will be embedded in another **said** tag, (either direct or free indirect)—see section 2.1.1 below. NB: The narrator of *To The Lighthouse* uses indirect discourse very rarely, so please consult with your TA before tagging such instances. Characters in *To the Lighthouse* do often introduce the thoughts or words of other characters through indirect discourse. Here is an example such embedded indirect discourse:

```
<said discourse="direct" …>"I overheard her say that <said
discourse="indirect" …>she adored it</said>"</said>
```

The third possible value for the *discourse* attribute is "free", which indicates free indirect discourse. Free indirect discourse is a mixture of narrative and direct discourse; like indirect discourse, it is in the third person and the past tense, but unlike indirect discourse, it contains words and expressions that indicate subjective or personalized aspects clearly distinct from the narrator's style. Otherwise stated, in free indirect discourse, 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 free indirect discourse, 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 free indirect discourse 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, quite often in *To the Lighthouse*, we are taken into a character's thoughts via free indirect discourse, and in these thoughts the words of another character are reported. For example,

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

NB: It can be difficult to distinguish between direct and indirect discourse within passages of free indirect discourse. You must simply do your best, based on the available evidence, to decide whether the character in whose head we have been placed via FID is attempting to render the *actual words* of another character's thoughts or speech (in which case, mark it as direct) or rather is more likely summarizing these thoughts or speech in his own words (in which case, mark it as indirect).

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 indirect discourse and in free indirect

discourse 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. A list of characters is given at the end of the blank TEI file. If the character speaking is not on that list, use "other". If multiple characters are involved, use "group" (you do not need to list each character in your tagging, but note that you DO need to explain your use of the "group" attribute in the written portion of Assignment #2). 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, however, that all text *not enclosed* by **said** tags is taken to be the narrator's.

NB: The **who** attribute is used to indicate the character *whose words are being reported/summarized* in the **said** span. The **who** attribute is NOT used indicate *who is reporting or summarizing these words*—that will already be clear from the context. In the following example, Billy is indicated in the **who** attribute as the speaker of the phrase about needing a break, even though Tammy paraphrases what he said in her own words. We already know from the tagging context that she is the one who introduces this passage of indirect discourse, and thus that the paraphrase of Billy is in her voice; there is thus no need to further indicate it.

```
<said discourse="free" who="#tammy" aloud="false">Tammy remembered how she'd
laughed her head off when Billy told her he <discourse="indirect" who="#billy"
aloud="true">needed to take a break from their relationship and find
himself</said>.</said>
```

3. THE CLUE TAG

The **clue** tag *must* be included within *every passage tagged as free indirect discourse*. It is used to indicate the specific words or expressions that lead the annotator to tag the passage as free indirect discourse. 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 much larger than **clue** spans; please do not make your entire **said** span a **clue** span simply to satisfy the condition that each span of free indirect discourse 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 Virginia Woolf's words appear, unaltered.

There is one instance in the assigned section of *To the Lighthouse* in which a character quotes a poem. Since these words belong not to another character within the storyworld of *To the Lighthouse*, but rather to a source that exists in the real world, it is not a candidate for the said tag. It has thus already been pre-tagged with the proper TEI element—the quote tag. Please leave this tag unaltered if you have been assigned the portion in which it appears.