What's in a name?

Rebecca Dridan

The Data

- ▶ Text taken from Linux blogs
- ▶ Automatic markup normalisation and sentence segmentation, manually corrected
- ▶ Parsed with off-the-shelf English Resource Grammar
- ▶ 452 manually annotated items
- ▶ 16% of errors in semantic arguments in compounds (compared to 9% PP attachment)
- ▶ 10/15 top (syntactic) label confusions relate to name vs. not-name distinctions

Domain-specific collocations

Examples of compounds mis-bracketed

(open source) software	(free software) community
(web server) header	(static IP) configuration
(small business) server	(file server) support

Domain-specific collocations

Examples of compounds mis-bracketed

(open source) software	(free software) community
(web server) header	(static IP) configuration
(small business) server	(file server) support
(Windows installation) CD	(FSF rights) definition

Domain-specific collocations

Examples of compounds mis-bracketed

(open source) software	(free software) community
(web server) header	(static IP) configuration
(small business) server	(file server) support

Windows (installation CD)   FSF (rights definition)

Domain-specific collocations

Examples of compounds mis-bracketed

(open source) software	(free software) community
(web server) header	(static IP) configuration
(small business) server	(file server) support

Windows (installation CD)   FSF (rights definition)

Solutions

- ▶ treebank large amounts of in-domain data
- ▶ use an external resource to enforce hard constraints
- ▶ learn domain-specific lexical model and integrate it with the parse ranking model

Domain-specific collocations

Examples of compounds mis-bracketed

(open source) software	(free software) community
(web server) header	(static IP) configuration
(small business) server	(file server) support

    Windows (installation CD)   FSF (rights definition)

Solutions

- ▶ treebank large amounts of in-domain data
- ▶ use an external resource to enforce hard constraints
- ▶ learn domain-specific lexical model and integrate it with the parse ranking model

However, only 10-15% of compound errors were of this type.

What's a name?

▶ Capitalisation triggers potential named entities.

I think Tree is coming.

Install Windows after Ubuntu.

What's a name?

- Capitalisation triggers potential named entities.

    I think Tree is coming.

    Install Windows after Ubuntu.

What's a name?

- ▶ Capitalisation triggers potential named entities.

    I think Tree is coming.

    Install Windows after Ubuntu.

- ▶ This can add unnecessary ambiguity to parsing.

    Beautiful Green Artwork

    New Package Management

What's a name?

▶ Capitalisation triggers potential named entities.

I think Tree is coming.

Install Windows after Ubuntu.

▶ This can add unnecessary ambiguity to parsing.

Beautiful Green Artwork

New Package Management

▶ And lead to difficult choices for the treebanker.

Our article entitled Ubuntu As Intended drew a fair amount of discussion.

What's a name?

▶ Capitalisation triggers potential named entities.

I think Tree is coming.

Install Windows after Ubuntu.

▶ This can add unnecessary ambiguity to parsing.

Beautiful Green Artwork

New Package Management

▶ And lead to difficult choices for the treebanker.

Our article entitled Ubuntu As Intended drew a fair amount of discussion.

What's in a name?

I have newly installed Ubuntu Lucid Lynx.

What's in a name?

I have newly installed Ubuntu *Lucid* *Lynx.*

What's in a name?

I have newly installed Ubuntu Lucid Lynx.

For Lucid Lynx users

What's in a name?

I have newly installed Ubuntu *Lucid* *Lynx.*

For *Lucid* *Lynx users*

For *Lucid* *users*

For *Lucid* *Users*

What's in a name?

I have newly installed Ubuntu Lucid Lynx.

For Lucid Lynx users

For Lucid users

For Lucid Users

Added support for NVIDIA 3D Vision Stereo.

Half Life 2 - Retail

The 195.22 NVIDIA Accelerated Linux Graphics Driver Set for Linux/x86 is available for download