



Too strait is the gate: tagging the aberrant in text-induced learner English

Robert Spence <mailto:spence@saar.de>

Fachrichtung 4.6

Universität des Saarlandes

PSF 15 11 50

D-66041

Germany

This PDF document will be available for download at:

<http://spence.saar.de/papers/liverpool2002.pdf>



A remark by a conference delegate:



2/17



A remark by a conference delegate:



2/17

'I can never understand your **titles** ...



A remark by a conference delegate:



2/17

‘I can never understand your **titles** ...

... and the **abstracts** don't really help much, either!’



A remark by a conference delegate:



‘I can never understand your **titles** ...

... and the **abstracts** don't really help much, either!’

— H. F., 14.07.2002



A remark by a conference delegate:



2/17

‘I can never understand your **titles** ...

... and the **abstracts** don't really help much, either!’

— H. F., 14.07.2002

... but maybe a laying bare of the **workpoints** might?



Workpoint 1:



Workpoint 1:



Enter ye in at the strait gate:
for wide *is* the gate,
and broad *is* the way,
that leadeth to destruction,
and many there be which go in thereat:

Because strait *is* the gate,
and narrow *is* the way,
which leadeth unto life,
and few there be that find it.



Workpoint 1: ...not a quote from 'Pulp Fiction'...



3/17

Enter ye in at the strait gate:
for wide *is* the gate,
and broad *is* the way,
that leadeth to destruction,
and many there be which go in thereat:

Because strait *is* the gate,
and narrow *is* the way,
which leadeth unto life,
and few there be that find it.



Workpoint 1: ...not a quote from 'Pulp Fiction'...



3/17

Enter ye in at the strait gate:
for wide *is* the gate,
and broad *is* the way,
that leadeth to destruction,
and many there be which go in thereat:

Because strait *is* the gate,
and narrow *is* the way,
which leadeth unto life,
and few there be that find it.

...but from the 'Sermon on the Mount' [Matthew 7: 13-14]



Workpoint 2: ...a 'gate'...



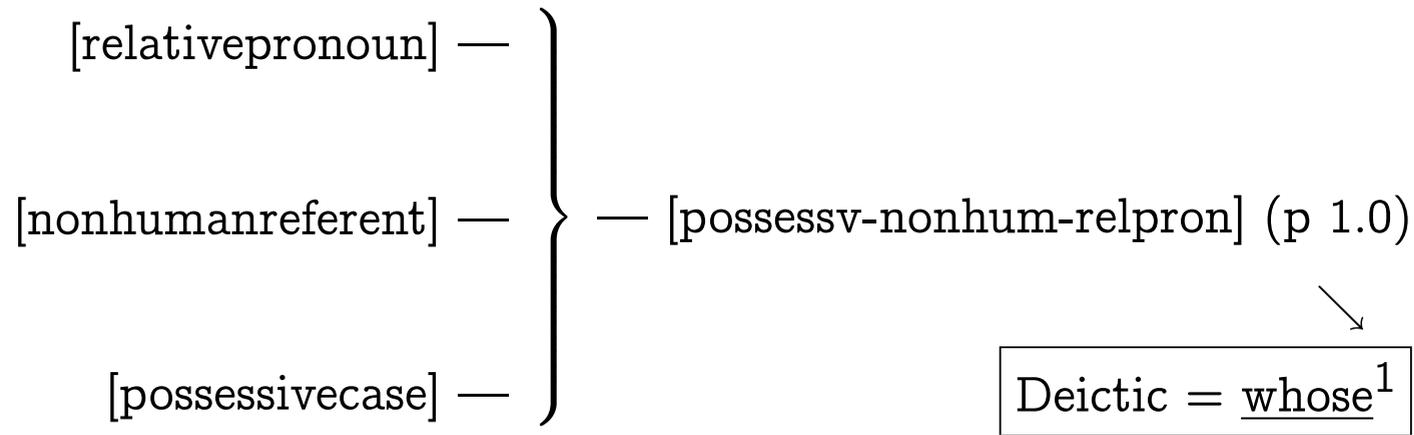
4/17



Workpoint 2: ...a 'gate'...



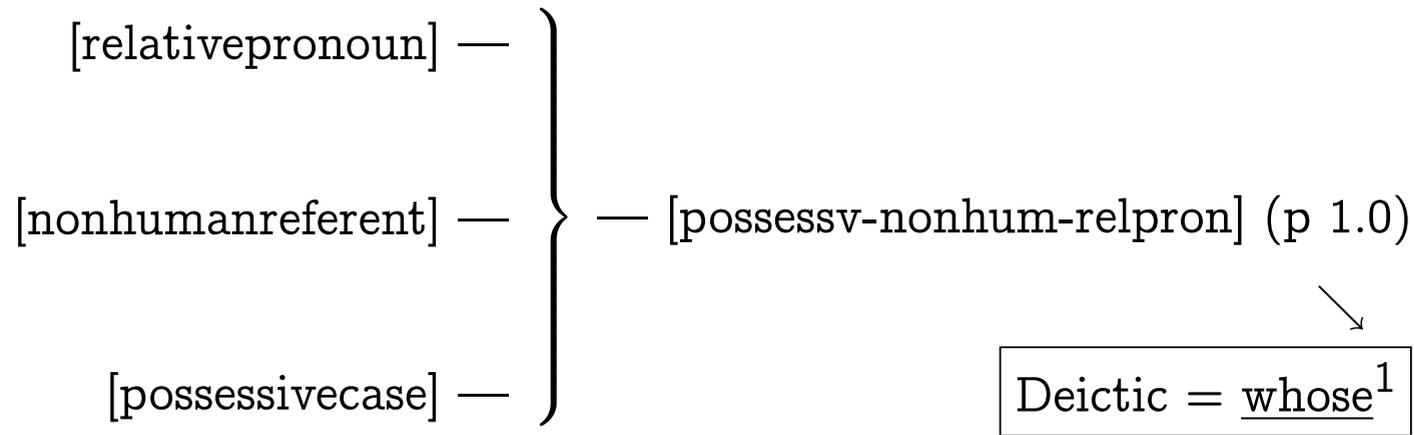
4/17



¹Example: This is an idea whose time has come.



Workpoint 2: ...a 'gate'...



¹Example: This is an idea whose time has come.

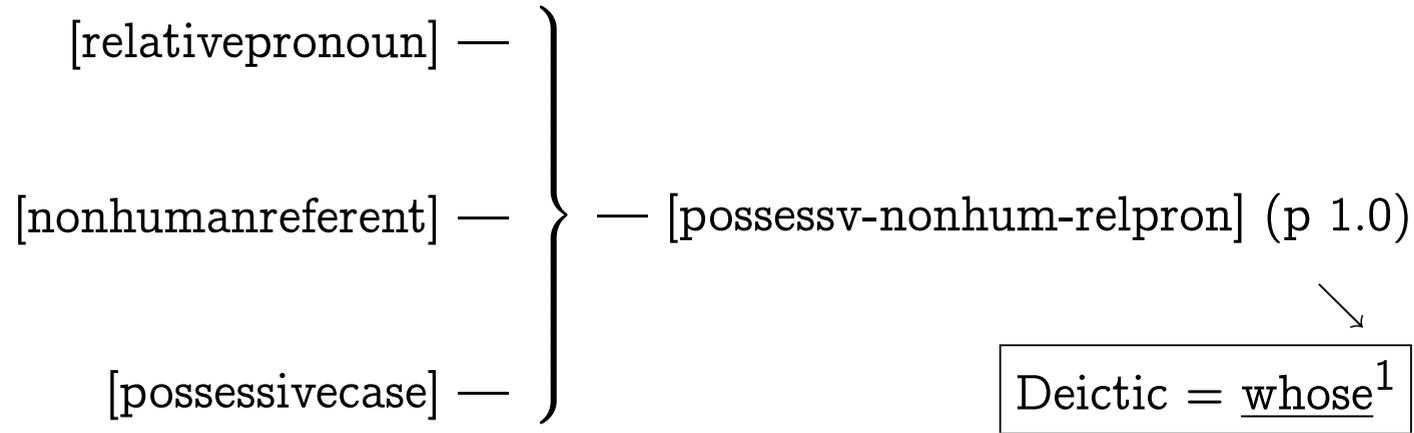
...a convention first used in the PENMAN NIGEL grammar



Workpoint 2: ...a 'gate'...



4/17



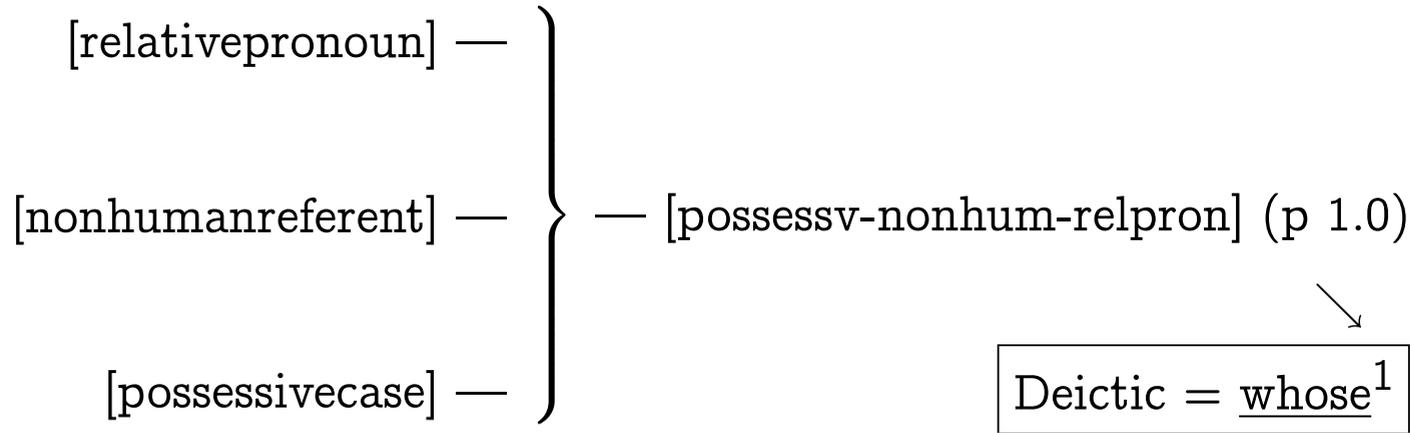
¹Example: This is an idea whose time has come.

...a convention first used in the PENMAN NIGEL grammar

...like a system, but with only one output



Workpoint 2: ...a 'gate'...



¹Example: This is an idea whose time has come.

...a convention first used in the PENMAN NIGEL grammar

...like a system, but with only one output

...a point at which systemic choice ends and automaticity of instantiation begins



Workpoint 3: ... 'text-induced' ...



Workpoint 3: ... 'text-induced' ...

Translation: 'text-induced text-production for a third party'
– Albrecht Neubert



Workpoint 3: ... 'text-induced' ...

Translation: 'text-induced text-production for a third party'

– Albrecht Neubert

Text production: activation of a semiotic potential



Workpoint 3: ... 'text-induced' ...

Translation: 'text-induced text-production for a third party'

– Albrecht Neubert

Text production: activation of a semiotic **potential**

In the case of L2 (L3, L4, ...) text production, the potential is often **non-standard**, and the product is often **aberrant**, 'wandering away' from the strait and narrow path of 'nativelikeness'



Workpoint 3: ... 'text-induced'...

Translation: 'text-induced text-production for a third party'

– Albrecht Neubert

Text production: activation of a semiotic **potential**

In the case of L2 (L3, L4, ...) text production, the potential is often **non-standard**, and the product is often **aberrant**, 'wandering away' from the strait and narrow path of 'nativelikeness'

L2 (L3, L4, ...) text production can be 'text-induced'...



Workpoint 3: ... 'text-induced' ...

Translation: 'text-induced text-production for a third party'

– Albrecht Neubert

Text production: activation of a semiotic potential

In the case of L2 (L3, L4, ...) text production, the potential is often non-standard, and the product is often aberrant, 'wandering away' from the strait and narrow path of 'nativelikeness'

L2 (L3, L4, ...) text production can be 'text-induced'...

... in the context of 'learning to translate'



Workpoint 3: ... 'text-induced' ...

Translation: 'text-induced text-production for a third party'

– Albrecht Neubert

Text production: activation of a semiotic potential

In the case of L2 (L3, L4, ...) text production, the potential is often non-standard, and the product is often aberrant, 'wandering away' from the strait and narrow path of 'nativelikeness'

L2 (L3, L4, ...) text production can be 'text-induced'...

... in the context of 'learning to translate'

... or in the context of 'proving one's command of L2 (L3, L4, ...)'



Workpoint 3: ... 'text-induced' ...

Translation: 'text-induced text-production for a third party'

– Albrecht Neubert

Text production: activation of a semiotic **potential**

In the case of L2 (L3, L4, ...) text production, the potential is often **non-standard**, and the product is often **aberrant**, 'wandering away' from the strait and narrow path of 'nativelikeness'

L2 (L3, L4, ...) text production can be 'text-induced'...

... in the context of 'learning to translate'

... or in the context of 'proving one's command of L2 (L3, L4, ...)'

How to **pin down** aberrant text-induced text production?



Workpoint 3: ... 'text-induced' ...



Translation: 'text-induced text-production for a third party'

– Albrecht Neubert

Text production: activation of a semiotic potential

In the case of L2 (L3, L4, ...) text production, the potential is often non-standard, and the product is often aberrant, 'wandering away' from the strait and narrow path of 'nativelikeness'

L2 (L3, L4, ...) text production can be 'text-induced'...

... in the context of 'learning to translate'

... or in the context of 'proving one's command of L2 (L3, L4, ...)'

How to pin down aberrant text-induced text production?

... in the system ('text-induced learner English')



Workpoint 3: ... 'text-induced' ...



5/17

Translation: 'text-induced text-production for a third party'

– Albrecht Neubert

Text production: activation of a semiotic potential

In the case of L2 (L3, L4, ...) text production, the potential is often non-standard, and the product is often aberrant, 'wandering away' from the strait and narrow path of 'nativelikeness'

L2 (L3, L4, ...) text production can be 'text-induced'...

... in the context of 'learning to translate'

... or in the context of 'proving one's command of L2 (L3, L4, ...)'

How to pin down aberrant text-induced text production?

... in the system ('text-induced learner English')

... and in the product (= in a corpus of translata evidencing the existence and properties of that variety of semiotic potential)?



Workpoint 4: ...in L2 English, often NOT a gate...



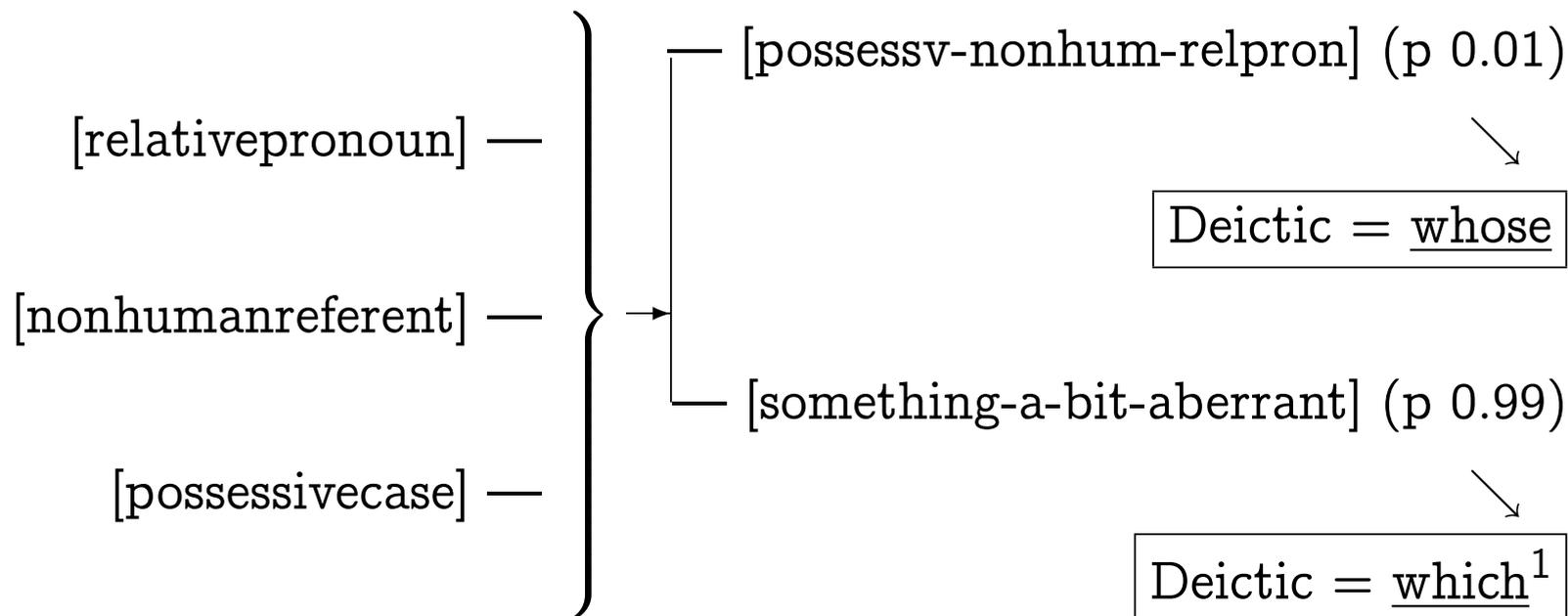
6/17



Workpoint 4: ...in L2 English, often NOT a gate...



6/17



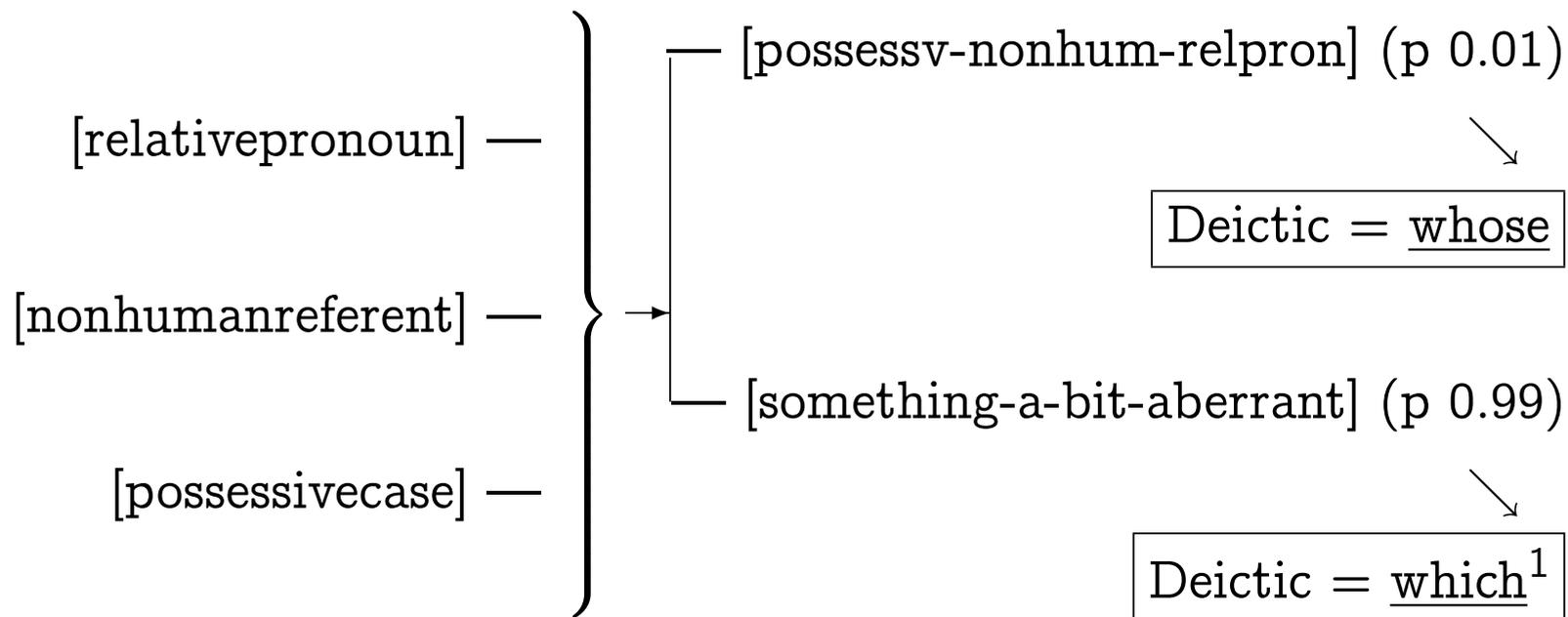
¹Example: This is an idea **which** time has come.



Workpoint 4: ...in L2 English, often NOT a gate...



6/17



¹Example: This is an idea **which** time has come.

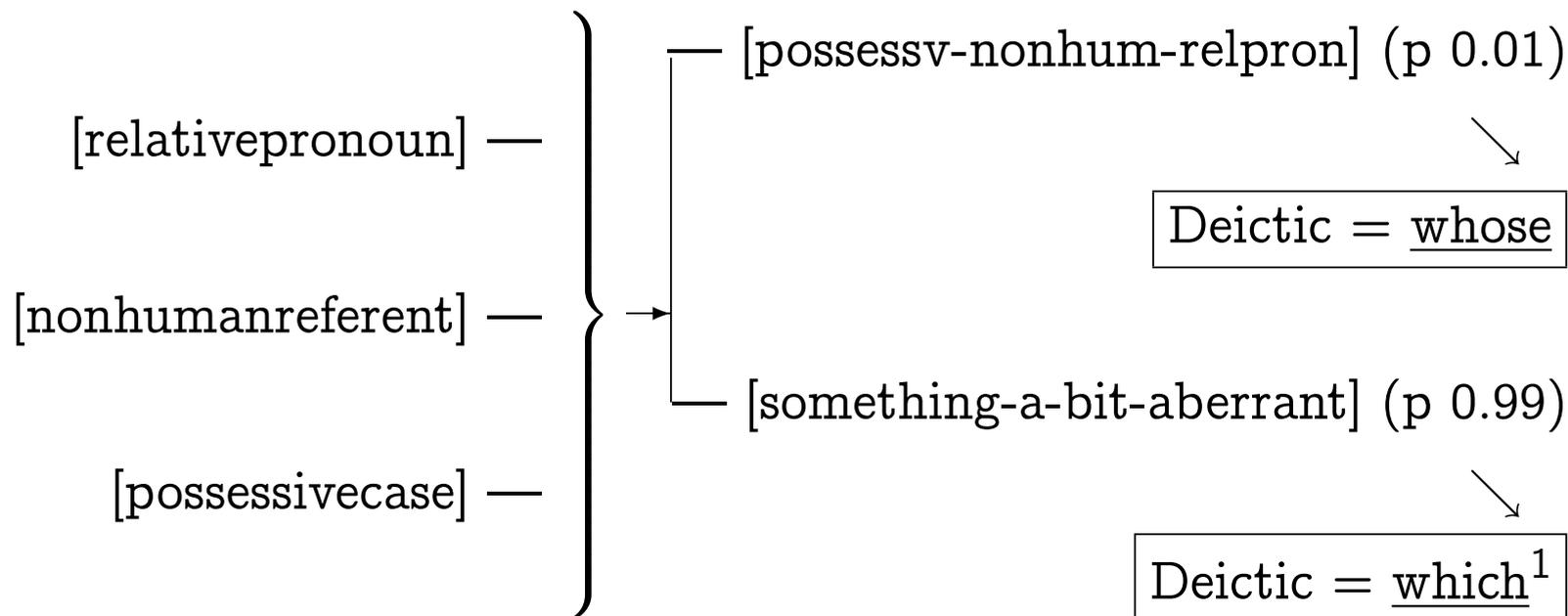
...so non-native aberrancy could be viewed (**through rose-tinted glasses**) as an expansion of the principle of systemic choice...



Workpoint 4: ...in L2 English, often NOT a gate...



6/17



¹Example: This is an idea which time has come.

...so non-native aberrancy could be viewed (through rose-tinted glasses) as an expansion of the principle of systemic choice...

...were it not for two problems:



How do we know what conjunction of features was to be realized?



7/17



How do we know what conjunction of features was to be realized?

...the role of the **L1 (/sourcelanguage) text** — inducing the activation of the L2 (/targetlanguage) text-production potential — can often provide a partial answer: ‘That’s the meaning that was in the source text — i.e. that’s the meaning that they **should have been** trying to translate’



How do we know what conjunction of features was to be realized?



7/17

...the role of the **L1 (/sourcelanguage) text** — inducing the activation of the L2 (/targetlanguage) text-production potential — can often provide a partial answer: ‘That’s the meaning that was in the source text — i.e. that’s the meaning that they **should have been** trying to translate’

...but is **[humanreferent]/[nonhumanreferent]** obligatorily coded in the source language at the stratum of **lexicogrammar**?



How do we know what conjunction of features was to be realized?



7/17

...the role of the **L1 (/sourcelanguage) text** — inducing the activation of the L2 (/targetlanguage) text-production potential — can often provide a partial answer: ‘That’s the meaning that was in the source text — i.e. that’s the meaning that they **should have been** trying to translate’

...but is **[humanreferent]/[nonhumanreferent]** obligatorily coded in the source language at the stratum of **lexicogrammar**?

How do we know it’s a system, not a gate?



How do we know what conjunction of features was to be realized?



7/17

...the role of the **L1 (/sourcelanguage) text** — inducing the activation of the L2 (/targetlanguage) text-production potential — can often provide a partial answer: ‘That’s the meaning that was in the source text — i.e. that’s the meaning that they **should have been** trying to translate’

...but is **[humanreferent]/[nonhumanreferent]** obligatorily coded in the source language at the stratum of **lexicogrammar**?

How do we know it’s a system, not a gate?

...‘that’s what’s in the corpus’



How do we know what conjunction of features was to be realized?



7/17

...the role of the **L1 (/sourcelanguage) text** — inducing the activation of the L2 (/targetlanguage) text-production potential — can often provide a partial answer: ‘That’s the meaning that was in the source text — i.e. that’s the meaning that they **should have been** trying to translate’

...but is **[humanreferent]/[nonhumanreferent]** obligatorily coded in the source language at the stratum of **lexicogrammar**?

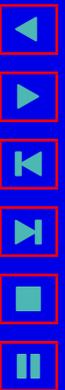
How do we know it’s a system, not a gate?

...‘that’s what’s in the corpus’

...the corpus is a manifestation of a transpersonal/collective semiotic potential

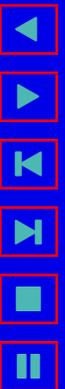


Some possible motivational and applicational contexts



Some possible motivational and applicational contexts

... more 'system-' or more 'product-' oriented...



Some possible motivational and applicational contexts

... more 'system-' or more 'product-' oriented...

... and/or more research- or teaching-oriented



Some possible motivational and applicational contexts

... more 'system-' or more 'product-' oriented...

... and/or more research- or teaching-oriented

1) **Praise or censure:** separation of trainee translators into 'better' and 'worse' ones



Some possible motivational and applicational contexts

... more 'system-' or more 'product-' oriented...

... and/or more research- or teaching-oriented

- 1) **Praise or censure:** separation of trainee translators into 'better' and 'worse' ones
- 2) **Diagnostics:** attempting to 'get at the root' of the evil



Some possible motivational and applicational contexts

... more 'system-' or more 'product-' oriented...

... and/or more research- or teaching-oriented

- 1) **Praise or censure:** separation of trainee translators into 'better' and 'worse' ones
- 2) **Diagnostics:** attempting to 'get at the root' of the evil
- 3) **Edification:** design of tools for guided self-improvement



Some possible motivational and applicational contexts

... more 'system-' or more 'product-' oriented...

... and/or more research- or teaching-oriented

- 1) **Praise or censure:** separation of trainee translators into 'better' and 'worse' ones
- 2) **Diagnostics:** attempting to 'get at the root' of the evil
- 3) **Edification:** design of tools for guided self-improvement
- 4) **Celebration of hybridization:** mapping ontogenetic/phylogenetic cross-fertilization



Some possible motivational and applicational contexts

... more 'system-' or more 'product-' oriented...

... and/or more research- or teaching-oriented

- 1) **Praise or censure:** separation of trainee translators into 'better' and 'worse' ones
- 2) **Diagnostics:** attempting to 'get at the root' of the evil
- 3) **Edification:** design of tools for guided self-improvement
- 4) **Celebration of hybridization:** mapping ontogenetic/phylogenetic cross-fertilization
- 5) **Automation:** meta-(cross-fertilization) with automatic error recognition systems



Some possible motivational and applicational contexts

... more 'system-' or more 'product-' oriented...

... and/or more research- or teaching-oriented

- 1) **Praise or censure:** separation of trainee translators into 'better' and 'worse' ones
 - 2) **Diagnostics:** attempting to 'get at the root' of the evil
 - 3) **Edification:** design of tools for guided self-improvement
 - 4) **Celebration of hybridization:** mapping ontogenetic/phylogenetic cross-fertilization
 - 5) **Automation:** meta-(cross-fertilization) with automatic error recognition systems
- here an 'errant' thought: is the potential behind a set of text-induced texts in any way similar to a **controlled language**?



Why XML?

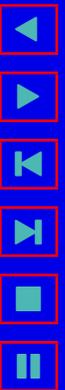


9/17



Why XML?

...because XML (Extensible Markup Language) is likely to remain — for some time — the standard for cross-platform distribution of ‘processable’ texts via the World Wide Web



Why XML?

...because XML (Extensible Markup Language) is likely to remain — for some time — the standard for cross-platform distribution of ‘processable’ texts via the World Wide Web

...because it’s an extensible metalanguage for defining one’s own customized markup languages in, not a single inflexible document type like HTML



Why XML?

...because XML (Extensible Markup Language) is likely to remain — for some time — the standard for cross-platform distribution of ‘processable’ texts via the World Wide Web

...because it’s an extensible metalanguage for defining one’s own customized markup languages in, not a single inflexible document type like HTML

...because it’s less complex than full SGML — i.e. it’s easier to write programs for in the cross-platform web-based programming language of one’s choice



An example



10/17



An example

German (=student's L1) source text (extract):

Im Jahre 1960 lebten z. B. auf dem afrikanischen Kontinent 276 Millionen Menschen;
inzwischen wuchs die Einwohnerzahl bis 1985 auf 553 Millionen
und hat sich damit in nur 25 Jahren mehr als verdoppelt.



An example



German (=student's L1) source text (extract):

Im Jahre 1960 lebten z. B. auf dem afrikanischen Kontinent 276 Millionen Menschen;
inzwischen wuchs die Einwohnerzahl bis 1985 auf 553 Millionen
und hat sich damit in nur 25 Jahren mehr als verdoppelt.

English (=student's L3) translation:

In 1960, for example, 276 million people lived on the African continent.
Till 1985 the population grew to 553 million.
This way it was more than doubling within 25 years only.



An example



German (=student's L1) source text (extract):

Im Jahre 1960 lebten z. B. auf dem afrikanischen Kontinent 276 Millionen Menschen;
inzwischen wuchs die Einwohnerzahl bis 1985 auf 553 Millionen
und hat sich damit in nur 25 Jahren mehr als verdoppelt.

English (=student's L3) translation:

In 1960, for example, 276 million people lived on the African continent.
Till 1985 the population grew to 553 million.
This way it was more than doubling within 25 years only.

How is this L3 target text extract aberrant?



An example



German (=student's L1) source text (extract):

Im Jahre 1960 lebten z. B. auf dem afrikanischen Kontinent 276 Millionen Menschen;
inzwischen wuchs die Einwohnerzahl bis 1985 auf 553 Millionen
und hat sich damit in nur 25 Jahren mehr als verdoppelt.

English (=student's L3) translation:

In 1960, for example, 276 million people lived on the African continent.
Till 1985 the population grew to 553 million.
This way it was more than doubling within 25 years only.

How is this L3 target text extract aberrant?

How aberrant is this L3 target text extract?





German (=student's L1) source text (extract):

Im Jahre 1960 lebten z. B. auf dem afrikanischen Kontinent 276 Millionen Menschen;
inzwischen wuchs die Einwohnerzahl bis 1985 auf 553 Millionen
und hat sich damit in nur 25 Jahren mehr als verdoppelt.

English (=student's L3) translation:

In 1960, for example, 276 million people lived on the African continent.
Till 1985 the population grew to 553 million.
This way it was more than doubling within 25 years only.

How is this L3 target text extract aberrant?

How aberrant is this L3 target text extract?



How the teacher might have 'marked' (= 'annotated' ? / = 'graded' ?)
the student('s (text (production (capability))))



How the teacher might have 'marked' (= 'annotated' ? / = 'graded' ?) the student('s (text (production (capability))))



12/17

In 1960, for example, 276 million people lived on the African continent.

Till¹ 1985 the population grew² to 553 million.

This way³ it was more than doubling⁴ within 25 years only⁵.



How the teacher might have 'marked' (= 'annotated' ? / = 'graded' ?) the student('s (text (production (capability))))



12/17

In 1960, for example, 276 million people lived on the African continent.

Till¹ 1985 the population grew² to 553 million.

This way³ it was more than doubling⁴ within 25 years only⁵.

Cross-referenced marginalia:



How the teacher might have 'marked' (= 'annotated' ? / = 'graded' ?) the student('s (text (production (capability))))



12/17

In 1960, for example, 276 million people lived on the African continent.

Till¹ 1985 the population grew² to 553 million.

This way³ it was more than doubling⁴ within 25 years only⁵.

Cross-referenced marginalia:

¹: prep



How the teacher might have 'marked' (= 'annotated' ? / = 'graded' ?) the student('s (text (production (capability))))



In 1960, for example, 276 million people lived on the African continent.

Till¹ 1985 the population grew² to 553 million.

This way³ it was more than doubling⁴ within 25 years only⁵.

Cross-referenced marginalia:

1: prep

2: T/A



How the teacher might have 'marked' (= 'annotated' ? / = 'graded' ?) the student('s (text (production (capability))))



In 1960, for example, 276 million people lived on the African continent.

Till¹ 1985 the population grew² to 553 million.

This way³ it was more than doubling⁴ within 25 years only⁵.

Cross-referenced marginalia:

1: prep

2: T/A

3: tr/conj/meaning



How the teacher might have 'marked' (= 'annotated' ? / = 'graded' ?) the student('s (text (production (capability))))



In 1960, for example, 276 million people lived on the African continent.

Till¹ 1985 the population grew² to 553 million.

This way³ it was more than doubling⁴ within 25 years only⁵.

Cross-referenced marginalia:

1: prep

2: T/A

3: tr/conj/meaning

4: T/A



How the teacher might have 'marked' (= 'annotated' ? / = 'graded' ?) the student('s (text (production (capability))))



In 1960, for example, 276 million people lived on the African continent.

Till¹ 1985 the population grew² to 553 million.

This way³ it was more than doubling⁴ within 25 years only⁵.

Cross-referenced marginalia:

1: prep

2: T/A

3: tr/conj/meaning

4: T/A

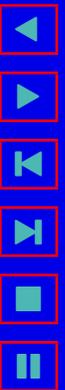
5: w.o.



How a human annotator might 'mark up' the text for corpus linguistic purposes: elements, tags, attributes



13/17



How a human annotator might 'mark up' the text for corpus linguistic purposes: elements, tags, attributes



13/17

In 1960, for example, 276 million people lived on the African continent.

<aberrancy> Till </aberrancy>

1985 the population

<dialect-dependent-possible-aberrancy>

grew

</dialect-dependent-possible-aberrancy>

to 553 million.

This way

<aberrancy

degree=serious

stratum-of-manifestation=lexicogrammar

probable-stratum-of-origin=lexicogrammar

rank-of-manifestation=group-phrase-rank

metafunction=textual:conjunctive-relation:cause-as-manner:over-

generalization-of-metaphorical-strategy:analytic-colloquial/>



it was more than doubling

<aberrancy

span-at-word-rank=-4

span-at-group-phrase-rank=-1 discontinuous/> within 25 years

only <aberrancy

degree=mild

stratum-of-manifestation=lexicogrammar

probable-stratum-of-origin=discourse-semantics

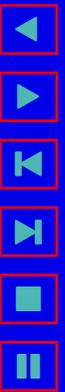
rank-of-manifestation=group-phrase-rank

metafunction=textual:addressee-oriented:culminative:

focus-on-semantic-restriction/>.



Don't let's forget, despite the discreetly frantic waving of reminder notices by the Chair...

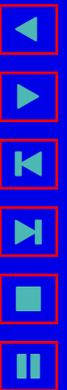


Don't let's forget, despite the discreetly frantic waving of reminder notices by the Chair...

The set of all possible attribute-value pairs for aberrancy tags can be understood as implying a system network.



15/17



Don't let's forget, despite the discreetly frantic waving of reminder notices by the Chair...



The set of all possible attribute-value pairs for aberrancy tags can be understood as implying a system network.

If lack of aberrancy and/or presence of geniality-of-transfer are additionally taken into account, such a network (or would it be a flowchart?) can be understood as a mapping of possible translation strategies.



Don't let's forget, despite the discreetly frantic waving of reminder notices by the Chair...



The set of all possible attribute-value pairs for aberrancy tags can be understood as implying a system network.

If lack of aberrancy and/or presence of geniality-of-transfer are additionally taken into account, such a network (or would it be a flowchart?) can be understood as a mapping of possible translation strategies.

The complexity of such a network would (if extended to the maximum possible degree of delicacy) exceed by far the complexity of either the source language text production potential or the target language text production potential.



Don't let's forget, despite the discreetly frantic waving of reminder notices by the Chair...



The set of all possible attribute-value pairs for aberrancy tags can be understood as implying a system network.

If lack of aberrancy and/or presence of geniality-of-transfer are additionally taken into account, such a network (or would it be a flowchart?) can be understood as a mapping of possible translation strategies.

The complexity of such a network would (if extended to the maximum possible degree of delicacy) exceed by far the complexity of either the source language text production potential or the target language text production potential.

Short-term feasibility of tagging automation is very low, as is likelihood of use in assessment contexts.



The degree of (censurable) aberrancy is relative to the number and nature of the instances of aberrancy as such, as weighted in accordance with the specification(s) of the translation task.



The degree of (censurable) aberrancy is relative to the number and nature of the instances of aberrancy as such, as weighted in accordance with the specification(s) of the translation task.

The degree of censure of recurrent instances of the same type of aberrancy is a function of the orientation of the observer — ‘practical usability of the target text for a specific purpose’ vs ‘psycholinguistic hypotheses about presumed conformity of competence (or about competence to conform)’.





The degree of (censurable) aberrancy is relative to the number and nature of the instances of aberrancy as such, as weighted in accordance with the specification(s) of the translation task.

The degree of censure of recurrent instances of the same type of aberrancy is a function of the orientation of the observer — ‘practical usability of the target text for a specific purpose’ vs ‘psycholinguistic hypotheses about presumed conformity of competence (or about competence to conform)’.

Aberrancy (hypothesized as originating) at higher strata is likely to involve longer spans with greater degrees of discontinuity of observed instantiation.



Experiment

Well this is just an experiment....



17/17

