



Automated Sports Journalism

Eero Laine



Sports journalism as content

- Game recaps
- Interviews
- In-depth analysis



Sports journalism as content

- Statistics feature prominently
- Uses game-related jargon
 - Eg. Grand Slam, Strikeout
- Uses stock phrases and lingo
 - Eg. "blowback", "comeback", "nail-biter"

A photograph of a baseball game in progress, showing a batter swinging and a catcher crouching. The image is dark and has a yellowish tint. The batter is in the center, wearing a white uniform with a '50' patch on the sleeve. The catcher is on the left, wearing a dark uniform and a mask. The background is a blurred crowd of spectators.

Game recaps

- An ideal format for automation
- Statistics as the core content
- Style and analysis minor factors
- Easy to generate

A photograph of a baseball game in progress, showing a batter swinging and a catcher crouching. The image is dark and has a yellowish tint. The text is overlaid on the image.

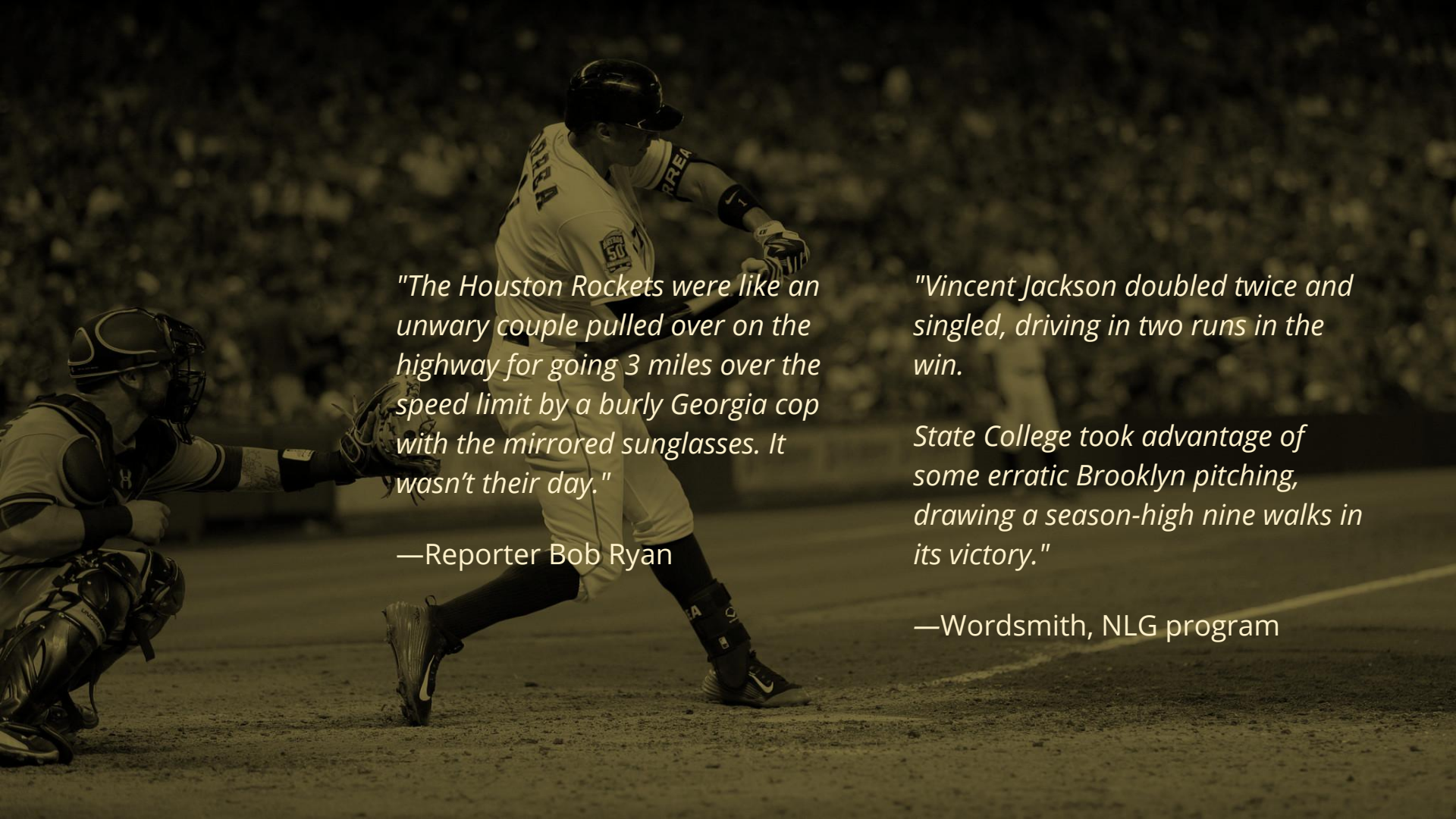
Game recap generation

- Focus on statistics-based facts
- Requires a feed of statistical data
- Data must be accurately submitted
- Available data limits content

A photograph of a baseball game in progress, showing a batter swinging and a catcher crouching. The image is overlaid with text and a list of features.

Features of automated game recaps

- Stock phrases repeated
- No personal style or analyses
- Formulaic



"The Houston Rockets were like an unwary couple pulled over on the highway for going 3 miles over the speed limit by a burly Georgia cop with the mirrored sunglasses. It wasn't their day."

—Reporter Bob Ryan

"Vincent Jackson doubled twice and singled, driving in two runs in the win."

State College took advantage of some erratic Brooklyn pitching, drawing a season-high nine walks in its victory."

—Wordsmith, NLG program

A photograph of a baseball game in progress, showing a batter swinging and a catcher crouching. The image is dark and has a greenish tint. The text is overlaid on the right side of the image.

Companies and products

- Automated Insights
 - Creator of Wordsmith
 - Formerly known as StatSheet
 - Notable customers: Gannett, AP
- Narrative Science
 - Creator of Quill
 - Background in the creation of StatsMonkey



Wordsmith

- Template-based
- Templates feature use of synonyms and conditional branches

Wordsmith UI example

✓ Done ↶ ↷ Insert Data Add Branch Add Synonym

Edit branch

Add conditional logic to your narrative:

- If this happens:
`sales > sales_previous`

Then write:
risen
- If this happens:
`sales < sales_previous`

Then write:
fallen
- + Add Condition

Or if none of the above are true, then write:
stayed the same

A photograph of a baseball game in progress. A batter in a white uniform is captured mid-swing, hitting the ball. A catcher in a dark uniform is crouched in the foreground, ready to catch the ball. The background shows a blurred crowd of spectators in a stadium. The overall image has a dark, moody tone with a yellowish-green tint.

StatsMonkey

- Created in Northwestern University
- Writes baseball stories from raw data
- Searches for key moments based on three baseball statistics
- Journalistic prioritization based on complex game-related algorithms



The Associated Press (AP) case

- NLG was employed to cover Minor League Baseball
- Automated Insights was hired by AP
- Increased coverage rather than replaced human reporters

A photograph of a baseball game in progress. A batter in a white uniform is captured mid-swing, hitting the ball. A catcher in a dark uniform is crouched in the foreground, ready to catch the ball. The background shows a large, blurred crowd of spectators in a stadium. The overall lighting is dim, suggesting an evening game.

Conclusion

- Automated sports journalism is focused on simple recaps
- Automated sports journalism does not have a high profile yet
- Automated Insights is currently the key company in the field